THREE ESSAYS IN APPLIED MICROECONOMICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ahmed Jaber

May 2014

THREE ESSAYS IN APPLIED MICROECONOMICS

Ahmed Jaber, Ph. D.

Cornell University 2014

This dissertation consists of three independent chapters that aim to further our understanding of the role of information on decision-making. The first chapter examines the effect of the introduction of broadband Internet on political behavior in the United States. I find that broadband has led to a large increase in voter turnout in presidential elections, as well as an increase in the total amount donated to political campaigns. Consistent with the hypothesis that Democrats have a stronger online presence, I also find a large effect on Democratic vote share in presidential elections. Evidence suggests the existence of both direct and indirect channels for the broadband effect. In particular, I find that broadband availability is associated with greater political knowledge, an increase in online (but not offline) donations, and the promotion of liberal values. In the second chapter, I examine the implications of a cognitive deficiency called "hindsight bias" on two legal standards: negligence and probable cause. When the court applies the negligence standard, I find that hindsight bias discourages (encourages) preventive measures that reduce the likelihood (severity) of an accident (respectively). When the court applies the probable cause standard, I show that hindsight bias discourages aggressive searches that increase the likelihood of police seizures. I also study the effectiveness of counterfactual thinking as a remedy for hindsight bias, and find that it is only unambiguously effective under the probable cause standard. In the third chapter, co-authored with Rebekka Christopoulou and Dean R. Lillard, we study smoking behavior in a model of cultural transmission. We derive conditions for the emergence and persistence of the smoking habit, and establish novel implications for the relationship between parental and societal influences. We then test and confirm the validity of our theory using novel data from the United States.

BIOGRAPHICAL SKETCH

Ahmed Jaber holds a B.A. in Mathematics and Economics from McGill University. After earning his degree, Ahmed worked as a full-time research assistant at the National Bureau of Economic Research in Cambridge, MA.

To Mom and Dad, for their endless acceptance.

To Sami and Line, for the missed moments.

To Muna, for her unfettered support.

To the pursuit of knowledge, for humbling my soul.

To the people of Syria, for my silence.

ACKNOWLEDGMENTS

I am grateful to Dan Benjamin, Aaron Bodoh-Creed, Steve Coate, Dean Lillard, Claire Lim, and Ted O'Donoghue for their complementary and non-substitutable advice. Dan Benjamin deserves a special mention for his continuous encouragement and tireless mentoring – mere words cannot do you justice. I also thank Rebekka Christopoulou, Dan Sacks and Russell Toth for extensive feedback on numerous drafts of the papers that make up this dissertation. The National Institute on Aging funded the research in Chapter 3 (through grant 1 R01 AG030379-01A2).

I would like to thank Samreen Malik and Max Mihm for being a family away from home – you gave meaning to the concept of Friendship. I thank Redouan Safdi for imparting with some of his knowledge – you embody the title that often precedes your name. I am grateful to numerous members of the Barazanji, Hamoush and Jaber families for their support throughout the years – your existence has shaped me in more ways than I realize. In particular, I would like to thank Hala Hamoush for more than living up to her role as an aunt.

TABLE OF CONTENTS

Chapter 1
Broadband Internet and Political Behavior: Evidence from the United States

1.1 Introduction

In 2008, Barack Obama led one of the most successful political campaigns in the history of the United States. His reliance on the Internet for mobilizing support and raising money has helped him earn the title of first Internet president. The 2008 presidential campaign took place at a time when more than half of all Americans were subscribed to broadband Internet. Would Obama's victory have been possible ten years earlier, when this technological revolution was not yet underway?[1] In this chapter, I formally investigate the effect of the introduction of broadband Internet on the U.S. political landscape.

Who did the Internet win over, and how? Even before Obama's historical campaign, Democrats used the Internet to attract attention to their cause. Howard Dean gained an early front-runner status in the 2004 presidential election by winning an "online primary" sponsored by MoveOn.org (Hindman 2005). The Internet decentralized the flow of information about politics, with the creation of new forms of political expression (e.g. blogs, Internet forums, social networking). Through websites, candidates were able to share information about campaign events with volunteers, and help organize local get-out-the-vote campaigns (Bimber and Davis 2003, Vaccari 2008). By combining positive and educational online messages, Democrats were also able to fight political disengagement (e.g. Kaid et al. 2007). These efforts were particularly successful among young college students with easy access to broadband. As a testament for these dynamics, the 2008 election was marked by a historically high turnout of the youth, who voted in a 2 to 1 ratio for Barack Obama.[2]

But while the Internet has facilitated the flow of information about politics, it has also brought with it an endless stream of entertainment options. Hence in principle, the opportunity cost of becoming informed could have risen, leading to political disengagement. Furthermore, the relevance

[1] SNL Kagan only estimates broadband availability starting in 1997, a time at which they estimate broadband to have reached 0.1% of households. By 1999, this number reaches only 1.5%. See http://tinyurl.com/krng8mg (retrieved on October 20, 2013).

[2] See Pew and Census reports at http://www.pewresearch.org/2008/11/13/young-voters-in-the-2008-election/ and http://www.census.gov/prod/2010pubs/p20-562.pdf (retrieved on November 12, 2013).

of information for voting behavior is still debated among theorists. For example, in the classical Downsian (1957) framework, voters base their vote on deeply rooted political views.

The effect of broadband Internet on political participation thus remains a fundamentally empirical question. To study the effect of broadband Internet, I use a panel of four presidential elections (every four years between 1996 and 2008), and zip code level panels of campaign donations for eight election cycles (every two years between 1994 and 2008). I include pre-broadband years as a reference point for the analysis, and end the study in 2008 for data availability reasons. The broadband measure is the number of Internet Service Providers, which I show is a good proxy for the number of broadband subscriptions.

The estimation of a broadband effect is complicated by the error in the measurement of broadband availability. It is also challenged by the existence of factors that correlate with both broadband availability and political participation. Using county / zip code fixed effects, I estimate the effect of the growth in broadband availability on the *change* in political participation. Since I only rely on variation from within a county / zip code, each geographic area is allowed to have its own baseline voting/donating behavior. To deal with the potentially confounding influence of time-varying characteristics, I also use an instrumental variable strategy based on land topography. The identifying assumption is that topography only affected the *change* in political participation in the 2000s – relative to pre-broadband years – through its effect on broadband deployment. I argue that low-lying areas are more prone to floods and exhibit higher summer temperatures. I then present evidence that these climate conditions hampered the deployment of broadband infrastructure. For example, flood protection required the erection of natural and artificial barriers to protect broadband facilities, while high temperatures made it necessary to use costly heat-resistant cable material.

To strengthen the validity of the identification strategy, I include state-time dummies as well as controls for zip code / county time-varying characteristics (e.g. income, population, education). This accounts for the potentially confounding influence of the electoral environment. I also include time dummies interacted with demographics that correlate with broadband availability. This allows areas that initially differ on observable characteristics (e.g. density, racial composition) to follow a different path in broadband deployment. Finally, I argue that the effect of broadband availability does not capture the influence of ongoing trends, e.g. through the influence of cable television.

Overall, I find that in presidential elections, broadband Internet has persuaded at least 29 percent of non-voters to cast a vote, and 10 percent of non-Democrats to switch sides. I compare the magnitude of these findings to the ones from the literature on persuasion, and find that broadband Internet has been a particularly persuasive medium of communication. I also estimate broadband explains 41 percent of the increase in campaign contributions between 1994 and 2008. However, this total effect on donations masks a great deal of heterogeneity across both donor and recipient characteristics. Dividing elections into presidential and mid-term cycles, I find that the effect of broadband is entirely driven by an increase in donations made in non-presidential years. I also find that in presidential election years, broadband caused a large re-allocation of funds toward Democratic presidential candidates, and away from House candidates and Political Action Committees. I then examine the mechanism behind the effect of broadband. I show that individuals living in areas with more broadband availability increased their consumption of online political news, donated more online (but not offline), exhibited greater levels of political knowledge, and were more likely to adhere to liberal values.

The findings on donating behavior indicate that the effect of broadband came from a reduction in the transaction cost of giving. As for the findings on voting behavior, they suggest that broadband has encouraged previously uninformed individuals to cast a vote. An effect of the media on political participation channeling through political knowledge has previously been documented for commercial TV (Prat and Strömberg 2005). More broadly, there is extensive theoretical and empirical evidence that information provision helps stimulate turnout (e.g. Matsusaka 1995, Feddersen and Pesendorfer 1996, Degan and Merlo 2009; Bartels 1996, Lassen 2005, Larcinese 2007, Battaglini et al. 2010).

Combined with the increase in turnout, the documented favorable vote swing toward Democrats suggests that the media can polarize individuals who have a preference for like-minded news (Mullainathan and Shleifer 2005, Sunstein 2007).[3] The association that I document between broadband and liberal values helps provide a mechanism for this polarizing effect. My findings suggest that broadband had an indirect effect on politics, by promoting social values that have been associated with the platform of the Democratic party. This argument is in the spirit of recent evidence that

[3]It is possible to generate a demand for like-minded news in both behavioral and rational models of behavior (e.g. Mullainathan and Shleifer 2005; Gentzkow and Shapiro 2006). Empirical evidence supports the existence of such "demand-driven" bias. For example, both Gentzkow and Shapiro (2010) and Puglisi and Snyder (2011) find that the ideology of U.S. newspapers conforms with the partisan composition of the geographic areas they circulate in.

television has indirectly favored Silvio Berlusconi by promoting a culture of individualism in Italy (Durante et al. 2013).

In a preliminary and independent study, Larcinese and Miner (2012) also consider the effect of the Internet on U.S. elections. They instrument for broadband availability using state-specific right-of-way laws. Larcinese and Miner find a positive effect of broadband on Democratic vote share, but cannot reject the absence of an effect on either turnout or campaign donations. They also do not attempt to uncover the mechanism behind the effect of broadband.

A few other contemporaneous studies attempt to disentangle the effect of broadband Internet on voting behavior in other countries, and find mixed evidence (Miner 2013, Czernich 2012, Falck et al. 2013, Campante et al. 2013).[4] My findings contribute to this literature in a number of ways. First, I document a robust positive effect of broadband on voter turnout in a well-established democracy. This effect on turnout lends credence to the idea that the Internet has fostered civic engagement. Second, I look at the effect of broadband on campaign donations, an important form of political participation that remains poorly understood (see Ansolabahere et al. 2003). Third, I suggest channels for the effect of broadband using data on political knowledge, political news consumption, moral values, and the medium for campaign donations (i.e. online or offline). Studying the mechanism behind the effect of broadband provides the necessary first step toward an understanding of why it differs so greatly across institutional contexts.[5]

My study also proposes a novel identification strategy to assess the effect of non-DSL broadband technologies. In contrast, the papers mentioned above identify the effect of DSL broadband Internet on political outcomes, by relying on measures of the distance between key components of the pre-existing telephony network (households, central office, backbone).[6] Since the dominance of DSL

[4]Miner (2013) studies Malaysian elections and finds that Internet penetration explains a third of the decrease in the incumbent party vote share. In an investigation of German elections, Czernich (2012) and Falck et al. (2013) find that broadband explains more than a third of the decrease in voter turnout in parliamentary elections in a subset of municipalities. Finally, Campante et al. (2013) find that broadband decreased voter turnout in Italian parliament elections between 1996 and 2008. But by 2013, this effect faded with the emergence of online grassroots movements. In the political science literature, researchers have also studied the effect of the Internet on voter turnout, online news consumption, and political attitudes. However, this evidence is mostly correlational, and has yielded mixed evidence (e.g. Bimber 2001, Tolbert and McNeal 2003, Golde and Nie 2010, Liang and Nordin 2012).

[5]To my knowledge, the study by Falck et al. (2013) is the only one that attempts to uncover the mechanism behind a broadband effect. As noted in footnote 4, their study argues for a negative effect of broadband on voter turnout. Using survey data, they then provide evidence that broadband crowded out television news, and led to a greater consumption of entertainment news. However, their definition of television news includes both political and non-political news, while their measure of entertainment is not limited to the online realm.

[6]The empirical approaches of the papers I have mentioned in this paragraph are all based on the structure of the pre-existing telephony network, which is only appropriate for understanding the effect of DSL broadband Internet. Czernich (2012) and

in the broadband market is slowly fading away, the identification strategy that I propose lays the building blocks for future studies of the effect of the Internet.[7] This potential is not limited to the realm of politics. In fact, numerous studies have argued for an effect of the Internet outside politics, including market competitiveness (Brown and Goolsbee 2002), international trade (Freund and Weinhod 2004), cigarette tax revenue (Goolsbee et al. 2010), female labor supply (Dettling 2013), sex crime (Bhuller et al. 2013), and marriage rates (Bellou 2013).

The rest of the chapter is organized as follows. Section 1.2 describes the data. Section 1.3 turns to the identification strategy. Section 1.4 contains the main analysis, along with robustness tests and placebo regressions. Section 1.5 studies the mechanism behind the documented effect. Section 1.6 concludes.

1.2 Data

Data on the availability of broadband Internet comes from the Federal Communications Commission (FCC). Every June 30^{th} and December 31^{st} between late-1999 and mid-2008, Internet Service Providers (ISPs) were required to report to the FCC all zip codes in which they had at least one high-speed Internet customer. The FCC defines a high-speed line as one that provides a connection speed exceeding 200 kb/s in at least one direction (upload or download). This includes Internet technologies like Cable Modem, DSL, and wireless.[8] In the rest of the analysis, I will use interchangeably the following expressions: “number of ISPs”, “broadband availability”, and “high-speed Internet availability”.

The main measure of broadband availability, $ISP_{z,t}$, is the number of ISPs operating in a zip code area z at time t.[9] The FCC lumps zip codes with 1, 2, or 3 ISPs into one category. I assume that “2” ISPs – i.e. the middle of the range – operate in each of these geographic areas. I then

Falck et al. (2013) instrument for broadband availability using the distance between a household and the closest central office. This factor determines whether it is possible to provide ADSL through traditional copper wires. Miner (2013) and Campante et al. (2013) use the distance between the central office and the backbone of the network as a basis for their instrumental variables. The underlying logic is that this distance affects the cost of laying out necessary fiber optic cables.

[7] For instance, the deployment of wireless Internet is more costly in forested areas like New England (Gillett et al. 2004). The use of terrain attributes is particularly promising given the availability of geological data at fine levels of geographic detail across the globe.

[8] Note however that dial-up connections are excluded, given their theoretical transfer speed of 56 kb/s. See footnote 50 for a discussion of why dial-up Internet is unlikely to have been an important determinant of political behavior.

[9] I also explore whether results are robust to an alternative measure of broadband availability. I define the extent of broadband diffusion as the number of ISPs operating in a zip code area weighted by the number of years they have been in the market. This modification to the raw measure $ISP_{z,t}$ is based on the idea that an ISP that has been operating for longer in an area

assign a value of "0" ISPs to zip codes where no high-speed line is reported (in a given year).[10] Because elections data is only available at the county-level, I also compute a county-level measure of the number of Internet Service Providers. I define broadband availability in a county as the population-weighted average of the number of ISPs in the zip codes contained in that county.[11]

This chapter therefore identifies the effect of the number of ISPs on political outcomes. However, the number of ISPs is only a measure of broadband availability if it exhibits a relationship with the share of broadband subscribers in a geographic area, which I refer to as "broadband penetration". Both theoretical and empirical evidence suggest such the existence of such a link, through the effect of competition among ISPs the price and quality of broadband service (e.g. Bresnahan and Reiss 1991; Aron and Burnstein 2003, Crandall et al 2004, Xiao and Orazem 2009).

Using county data from the FCC, I also provide empirical support for the relationship between broadband availability and penetration. In particular, I examine the relationship between the number of ISPs and the percentage of broadband households. Unfortunately, the broadband availability and penetration data series do not overlap; the former data series *ends* in June 2008, while the latter *starts* in December 2008. To maximize the comparability of these data, I compare the 2008 time series for these two variables.

In Table 1.1, I report estimates from regressions of broadband penetration on broadband availability. These regressions always include data from 2008, but either include or exclude data from

may have achieved greater market penetration. Formally, I construct diffusion in the following way:

$$\text{diffusion}_{z,t} = \Sigma_{\tilde{t}\leq t} \left(\frac{t-\tilde{t}}{2}\right) \cdot ISP\_new_{z,\tilde{t}},$$

where z denotes a zip code area, $\left(t,\tilde{t}\right) \in \{\text{June 1996, December 1996}, ..., t\}^2$ indicate half-years, and $ISP\_new_{z,\tilde{t}} \equiv ISP_{z,\tilde{t}-1} - ISP_{z,\tilde{t}}$ denotes the difference in the number of ISPs between time $\tilde{t}$ and time $\tilde{t}-1$. There are three implicit assumptions that go into the construction of this measure of broadband diffusion. First, the growth in market penetration of an ISP grows linearly with time. Second, new ISPs start operating halfway through the semester being considered (factor of "2" in the denominator). Third, entry / exit into a zip code broadband market can be captured through the evolution in the *number* of ISPs through time (the FCC data do not report the name of the ISPs included in the zip code-level counts). For all practical purposes, using this diffusion leads similar broadband effects to the ones obtained in the analysis reported in the body of the chapter (both qualitatively and quantitatively).

[10] This does not indicate that there is exactly zero broadband subscriptions. In fact, broadband providers were initially only required to report their activity in a given state only if they had a minimum of 250 lines. This minimum reporting requirement was however dropped in December 2004, and this had little effect of this change on zip code level aggregates of the number of ISPs (c.f. Table 1.15 and Chart 12 in FCC, 2009).

[11] Formally, the amount of broadband availability in county c at time t is $ISP_{c,t} \equiv \frac{\sum_{z_c\in c} ISP_{z_c,t}\cdot p_{z_c}}{\sum_{z_c \in c} p_{z_c}}$, where the subscript z_c indicates a zip code contained within county c, and p_{z_c} denotes the 2000 Census population of that zip code. When zip codes are not fully contained within counties, I allocate the population of each zip code equally over all counties it is contained in. I do not limit myself to zip codes that are fully contained in counties, as this would force me to drop a large – and non-random – sample of counties from the analysis.

1996 and county dummies. In the first set of regressions, the same high-speed lines are included under the definition of both broadband penetration and broadband availability (connection speeds of 200 kb/s in at least one direction). In the second set of regressions, I base broadband penetration on a more conservative definition, which only includes internet lines with download speeds of at least 768 kb/s.[12] This alternative measure became the FCC's official definition of broadband starting in December 2008.[13] It has since been featured in policy debates, like the American Recovery and Reinvestment Act of 2009.

In the regressions which utilize data from 1996 and 2008, an additional ISP predicts an increase in broadband penetration of at least 4.38 percent. Relying only on data from 2008, this correlation between broadband availability and broadband penetration reduces to at most 3.15 percent. In all cases, estimates are highly statistically significant, which suggests that broadband availability is a good proxy for broadband penetration. However, regressions utilizing data from both 1996 and 2008 explain more variation in broadband penetration than those relying solely on 2008 data ($R^2 \geq 0.71$ vs. $R^2 \leq 0.174$, respectively). Therefore, broadband availability is a more appropriate proxy for broadband penetration in an investigation that relies on within-county variation, rather than cross-county variation.

To assess the effect of broadband on presidential election outcomes, I match county-level broadband data with data on presidential votes. Data on presidential election returns come from the CQ Press Voting and Elections database. These data, whose time frame extends from 1980 to 2008, include for each county the number of votes cast and their distribution across candidates. I define voter turnout as the proportion of the adult voting-age population who cast a ballot.[14] I then focus on vote shares for the two major parties in the U.S. political system, the Democratic party and the Republican party. The county elections dataset is constructed in a straightforward way. I only exclude 3 counties for which I do not have election data for all years. This leaves me with 3,105 counties and county-equivalents in the 48 contiguous states.[15]

[12] The upstream speed requirement remains 200 kb/s under this alternative definition of broadband.

[13] See statement of Chairman Kevin J. Martin at http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-280909A2.doc (retrieved on October 20, 2013).

[14] An alternative definition of voter turnout is relative to the number of registered voters. The literature has however avoided using it, for reasons pointed out by Gentzkow (2006) quoting Prysby (1987): "Calculating turnout as a proportion of registered voters [...] is generally inappropriate in the American context, given the large numbers of people who are eligible to register but fail to do so. In fact, the available research indicates that most nonvoters are not registered."

[15] The U.S. federal government uses the term "county-equivalent" to describe Louisiana parishes; and the independent cities of the states of Virginia, Maryland, Missouri, and Nevada that are not a part of any county, nor considered a consolidated

To understand the effect of broadband on campaign contributions, I combine zip code-level broadband data with data on campaign contributions from the Federal Elections Commission (FEC).[16] For each monetary donation, the FEC collects data on the characteristics of both contributors and recipients. These data allow a study of the effect of broadband on campaign donations to presidential candidates, Congressional candidates, and Political Action Committees (while the voting data only reflects participation in presidential elections). In the main analysis, I aggregate data on campaign donations at the zip code / election cycle level. In an ancillary analysis, I investigate the potentially heterogeneous effect of broadband by constructing separate panels on each donor / recipient characteristic (e.g. gender of contributor, political party of recipient).

The number and boundaries of zip codes change through time, and some zip codes are dedicated for military purposes. To avoid artificially assigning zero donations to some zip codes, data on donations is only included for zip codes for which the US Census reports a population estimate in 2000. For the 48 contiguous United States, this leaves me with a sample of 27,405 zip codes, which house 99.1 percent of the U.S. population.[17] For each one of those zip codes, I only include a portion of all donations. This is partly due to data limitations, as only money given by donors who contributed at least $200 in an election cycle are included (FEC reporting requirement).[18] I also voluntarily only include campaign contributions emanating from individuals not running for office.[19] Overall, this is a small limitation as the donations that remain in the dataset constitute the bulk of the money that goes toward funding political campaigns (e.g. Ansolabahere et al., 2003).

city–county.

[16] The Center for Responsive Politics assembles into a single database all forms filed by donors to the FEC. To minimize the extent of measurement error, I perform some data cleaning on this dataset. First, I delete duplicate transactions, e.g. same amount, individual, date, and recipient (3.3 percent of transactions in 2008). Second, I cap donations at the legal contribution limits, which are described in detail in Appendix Table A1 (0.21 percent of transactions in 2008).

[17] Since 1963, the United States Postal Service (USPS) has organized mail delivery by dividing geographic areas into zip codes. Keeping with this original intention, the USPS periodically updates the boundaries of zip codes, and new zip codes are added to the list of existing ones. (e.g. see http://about.usps.com/who-we-are/postal-facts/; retrieved November 29, 2013). I obtain a classification of zip codes into four "classes" : "Standard," "Military," "PO Box" and "Unique"; see http://www.populardata.com (retrieved on October 20, 2013). The analysis only includes "standard" zip codes, which are likely to refer to primary living headquarters.

[18] The majority of individual donations come from donors giving more than $200. This observation is true even for the 2008 Obama presidential campaign. Despite being celebrated for its reliance on small donor money, Obama raised only about a quarter of all its funds through donations below the $200 threshold. See ABC News article at http://abcnews.go.com/Politics/story?id=6329217&page=1 (retrieved on October 20, 2013).

[19] I exclude candidates' self-financing, the only form of fundraising that is not capped by the FEC (see Appendix Table A1 for legal limits). This helps prevent generating misleading results due to outliers. I also exclude donations made by Political Action Committees (PACs), since it is difficult to assign a meaningful measure of broadband availability to them. In particular, their operations are often spread out through geographic space. It would thus not be reasonable to base the broadband availability on, say, the zip code or county its headquarters are located in.

To identify the effect of broadband on political outcomes, I construct two county-level instrumental variables based on land altitude using ArcGIS. Data on land elevation is available for each 90-square meter of the globe, but I convert them to county-level means to match them with the presidential elections and demographics panels described above. In particular, I use data on land elevation collected in 1996 by the U.S. Geological Survey (USGS). Using the land elevation data, a measure of flow accumulation is also constructed to capture the location of where rainfall water accumulates.[20]

To study the mechanism behind the effect of broadband, I merge broadband data with individual-level data from the Pew Research Center. Over the course of the past two decades, Pew has fielded numerous surveys on themes related to political behavior and Internet use.[21] I include data from Pew surveys that measure respondents' political knowledge, moral values, preferred media for political news (e.g. radio, internet), and medium for campaign donations (i.e. online or offline). I further limit myself to surveys fielded in 2000 or later, as earlier surveys do not provide detailed data for where the respondent lives (either county or zip code). Without detailed geographic data, I cannot match each survey participant with the ISPs measure described above, and therefore cannot identify the broadband available to them.

1.3 Empirical strategy

1.3.1 The introduction and growth of broadband availability

Market analysis suggests there has been strong demand for broadband connectivity since the early 2000s (Faulhaber 2002). However, technological limitations and cost considerations have made broadband availability lag behind. At the core, the major impediments to broadband growth are all related to the structure of cable companies' networks. Traditional cable networks were only designed to transmit one-way television signals. In order to provide broadband, cable companies had to make their networks compatible for two-way data transmission. In practice, cable firms chose to upgrade to high-bandwidth Hybrid Coaxial-Fiber (HCF) networks, which can transmit higher-quality data (see Figure 1.1).

[20] Flow accumulation is a derivative measure based on land elevation. For any given 90-square meter cell in a grid, it is defined as the number of cells that lie at a higher level of elevation.

[21] All the information relevant to Pew surveys is available online at http://www.pewresearch.org/

The upgrade to HCF networks was limited in its scope. Firms only had the incentive to enter markets where they expected to recoup an appropriate rate of return on their investment (Faulhaber 2002). Because a large portion of the expense came from the fixed cost of connecting a neighborhood to the backbone of the network, cable companies only provided broadband to areas with at least 500-2000 customers (e.g. Hazlett and Bittlingmayer 2003). The upgrade to broadband connectivity was also not instantaneous. An important impediment to speedy broadband growth was the heavy construction work required to lay out new cables. From cable firms' perspective, the early deployment of infrastructure was also undesirable since broadband systems are more expensive to maintain than traditional cable networks. Furthermore, cable companies are likely to have been facing severe liquidity constraints. In fact, the cost of deploying broadband infrastructure is prohibitive, averaging $200 per home passed.

Figure 1.2 give a visual representation of the broadband market, through maps depicting broadband availability at the beginning of, and half-way through, the sample period (December 1999, December 2003). There is clearly a great extent of both within- and between-zip code variation in broadband availability. Table 1.2 then summarizes the distribution of broadband availability over seven election cycles. In 1996, no broadband was available to U.S. customers. Overall, broadband availability steadily increases through time. In June 2008, the median number of ISPs was 7 (50 percent broadband penetration), and only 55 zip codes (i.e. 0.2 percent of all zip codes) had no broadband available.[22]

Cable infrastructure is subject to damage from flooding, high ground temperatures, and excessive precipitation (Zimmerman and Faris 2010). There is strong evidence that climate conditions affect both the cost and time required to construct broadband infrastructure. For example, these challenges from nature require the use of damage-resistant material and the construction of natural / artificial barriers (Rosenzweig et al. 2011). They also generate larger maintenance costs once the infrastructure has been constructed. In this chapter, I argue that by capturing part of the influence of climate, lower land elevation correlates positively with the cost of building infrastructure.

The reliance on land elevation to capture a multitude of influences at once is not unconventional. In fact, geologists have shown that land elevation predicts soil characteristics such as water content,

[22] Given that I set to "0" the number of ISPs for zip codes absent from the FCC data in a given year, this is a strong indication that none of the zip codes in my sample are systematically excluded from the FCC broadband data.

crop yields, and vegetation type (Moore et al. 1991, Erskine et al. 2007). Moreover, the availability of elevation data at fine levels of aggregation is itself testimony for its predictive power. It is not a coincidence that the USGS has devoted considerable time and effort constructing elevation maps for the entire globe.

Low-lying areas are subject to greater flooding of telecommunications infrastructure through storm surges and hurricanes (e.g. Michel-Kerjan et al. 2010, Landry and Parvar 2011). When facing flood risk, broadband providers must invest more to safeguard their broadband facilities from being submerged under water (e.g. Nordhaus 2010, Rosenzweig et al. 2011).[23] For example, following Hurricane Sandy back-up generators failed in Lower Manhattan's broadband facilities, which led to the shut down of the telecommunications network.[24] Another challenge associated with being in a flood-prone area comes from the difficulty of burying cable underground. In turn, this may affect the cost of building broadband infrastructure, as overhead poles are in general more susceptible to weather damage (Bascom and Antoniello 2011).[25]

There also exists a strong negative relationship between elevation and summer temperatures (e.g. Willmott and Matsuura, 1995).[26] Evidence suggests that high temperatures damage cables carrying too much power. This is because cables produce heat, which needs to be dissipated in the ground.[27] When faced with this challenge, broadband providers incur higher installation and maintenance costs (e.g. additional cables, or artificial soil to absorb heat).

The many challenges I have outlined are heavily associated with the use of cable technology. This makes them unique to the United States, where cable companies have controlled the lion share of the broadband market (85 percent in 2003). In contrast, most other industrialized nations have relied on ADSL broadband technology (e.g. Western Europe). Unlike cable technology, ADSL requires minimal investment in infrastructure to be able to carry broadband signals. The conscientious

[23] In inland areas, flood damage comes from the accumulation of rainfall water; see NYC's report http://tinyurl.com/mtyrhem (retrieved on October 20, 2013). In coastal areas, flood damage comes from water corrosion due to sea level rise.

[24] See also NYC's report at http://tinyurl.com/mtyrhem (retrieved on October 20, 2013).

[25] There are advantages to each cabling approach. For example, underground are convenient to use in densely populated areas and to cross rivers, less prone to some types of damage (e.g. severe weather, theft), have a lower environmental impact (e.g. emission of electromagnetic fields), and need less clear surrounding area (e.g. airplanes). As for overhead cables, they are more accessible when damaged (fault location, management of dielectric oil in cables), less subject to some types of damage (e.g. unaware digger, earthquakes), cheaper to install, and requires less difficult voltage control. See http://tinyurl.com/ao4u4pc, http://tinyurl.com/m4rwx78 and http://tinyurl.com/mtlnwra (all retrieved on October 20, 2013).

[26] This relationship depends heavily on the atmosphere being well-mixed, which means it holds particularly well in inland areas. The main summer exception is in coastal regions with well-defined marine layers, where maximum temperatures often increase with elevation above the marine inversion (Daly et al. 2008).

[27] See http://www.atc-projects.com/learning-center/underground-transmission-lines/ (retrieved on October 20, 2013).

reader may then wonder why ADSL has not played a more major role in the U.S. context. I turn to this issue in some detail in the Appendix, where I discuss the market, technological, and regulatory limitations that have hampered ADSL provision in the United States.

1.3.2 Identification strategy

The goal of this chapter is to estimate if broadband availability affects voting patterns in presidential elections, and the amount donated to federal political causes (i.e. Congressional / Presidential candidates, PACs). There has been much discussion of the potential effect of the Internet on U.S. politics. However, identifying its causal effect is a challenging task. As previously mentioned, the introduction and growth of broadband Internet was far from uniform across the United States. To the extent that some of this variation reflects the different underlying demand of these counties, any correlation between broadband availability and political outcomes would have to be interpreted with caution. In particular, the unobserved characteristics of the residents of a geographic area may affect both broadband and political outcomes. For example, broadband providers found it more profitable to provide broadband to wealthier and more educated geographic areas. In turn, these areas have exhibited different trends in political participation over time. A raw correlation between broadband availability and political participation may then reflect the influence of such omitted variables from the model.

A second challenge to identification comes from the fact that broadband availability is measured with error. If it is classical, this measurement error in my data on broadband availability will cause attenuation bias. First, data limitations force me to manually assign a single value for broadband availability to zip codes where 1 to 3 ISPs operate. This is an important limitation, as it applies to 41.2% of zip code broadband data assigned to the five election cycles between 2000 and 2008. Second, I am forced to introduce error because the data on voting patterns and ISPs are partially geographically mismatched.

It is informative to pinpoint county characteristics that have a bearing on both broadband availability and political participation. Table 1.3 presents socio-economic characteristics of geographic areas by broadband status. I compute the summary statistics of zip code areas and counties for the top and bottom half on broadband availability in 2000. On average, I find that geographic areas with greater broadband availability are more densely populated, more urban, wealthier, more educated, and younger. Importantly, high broadband areas also contribute more money to campaigns, and exhibit lower voter turnout.

A simple regression makes more precise the key characteristics which determine broadband deployment (regression estimates not reported). Using the 2000 data cross-section, I find that the number of ISPs correlates strongly with median age, income, federal expenditures, population, log farm value, % black, % below poverty line, % high school grads, % income \$100k+, % urban, and % pop 18+ and % pop 65+ ($R^2 = 0.56$).[28] However, when one regresses the residuals from the previous regression on the other demographics listed in Table 1.3, the coefficients are non-statistically distinguishable from zero. This suggests that the remaining variation is largely idiosyncratic. Adding state fixed effects, however, helps capture some of the remaining variation ($R^2 = 0.61$).

1.3.3 Empirical specification

One of the contributions of this study is to provide a framework for the study of the effect of broadband that mitigates the potential bias from omitted variables and measurement error. To do so, I rely on panels of zip code and county level data. I then combine a fixed effects regression framework with an instrumental variables strategy. Formally, I estimate the following system of equations:

$$\begin{aligned} \mathrm{Y}_{it} &= \alpha_i + \delta_{st} + \gamma \mathbf{X}_{it} + \beta \mathrm{ISP}_{it} + \varepsilon_{it}, \qquad (1) \\ \mathrm{ISP}_{it} &= \widetilde{\alpha}_i + \widetilde{\delta}_{st} + \widetilde{\gamma} \mathbf{X}_{it} + \eta \mathrm{Z}_{it} + \nu_{it}, \qquad (2) \end{aligned}$$

The subscript i denotes a "geographic area", which should be interpreted as a zip code or county depending on the context. The subscript t refers to a presidential election date (e.g. November 4 2008) or campaign contributions cycle (e.g. January 1^{st} 2007 to December 31^{st} 2008). The variable ISP_{it} denotes the number of Internet Service Providers in geographic area i at time t.[29] The political outcome Y_{it} denotes campaign donations, voter turnout, or Democratic vote share. Finally, the specification includes a number of controls at different levels of aggregations: zip code

[28] The Census defines as "urban" census blocks that are densely populated (≥1,000 people per square mile), and that are surrounded by highly dense census blocks (≥500 people per square mile). See http://www.census.gov/geo/reference/ua/urban-rural-2000.html (retrieved on November 28, 2013)

[29] Data on the number of ISPs is only available up until June 2008, and so I use this early value of diffusion to capture broadband availability in the 2007-2008 election cycle.

or county fixed effects (α_i and $\widetilde{\alpha}_i$), depending on whether the analysis is for presidential elections or campaign contributions (respectively), state-time fixed effects (δ_{st} and $\widetilde{\delta}_{st}$) and time-varying area-level characteristics ($\mathbf{X}_{it}$).

Equation (1) models a political outcome Y_{it} as a function of an an endogenous regressor ISP_{it}, Equation (2) is the corresponding first-stage equation where Z_i, a measure of land elevation described below, serves as an excluded instrument. Geographic areas are defined at the lowest possible level of data aggregation for the dependent variable Y_{it}. This corresponds to the zip code level for campaign donations, and to the county level for presidential voting outcomes.

I estimate the effect of broadband using zip code variation in the analysis of campaign donations, and county variation in the analysis of presidential elections voting. I thus compare the growth of broadband availability with the evolution of different forms of political participation. Any relationship that is a function of the baseline characteristics of zip codes / counties would be absorbed by the zip code / county dummies. The instrumental variable strategy, which is inspired by the effect of topography on the cost of building broadband infrastructure, then deals with the potential influence of time-varying factors. Since topography does not vary in any significant way over the course of a few decades, I cannot use time variation as the source for identification. Instead, the estimation strategy instead relies on the relationship between *time-invariant* topographic measures and broadband deployment.

The main instrumental variable used in the analysis is $\log(\text{mean elevation})_i \cdot I_{t\geq 2000}$, i.e. a measure of altitude interacted with a post-broadband dummy.[30] The use of a single post-broadband dummy identifies the average effect of land elevation across broadband years.[31] The identifying assumption is that land elevation causes a change in political outcomes between pre- and post-broadband years only through its effect on broadband deployment. To be clear, the use of zip code- / county-level fixed effects does not prevent the political culture to systematically differ in more hilly areas, or topography to affect the rate of change in political participation *before* 2000.

The state-time fixed effects (δ_{st} and $\widetilde{\delta}_{st}$) account for the differing electoral stakes and broadband

[30] In order to more directly capture the influence of rainfall accumulation, I also attempt to include an additional instrument, $\log(\text{flow accumulation}) \cdot I_{t\geq 2000}$. Because the first-stage of this potential instrument is weak, I however drop it from the main analysis. Results are however similar when both instruments are included. Findings are in fact slightly larger in magnitude, which reinforces the idea that attenuation bias reduces the magnitude of the raw correlation between broadband and political participation

[31] In Section 1.4.1, I test empirically the hypothesis that the effect of land elevation on broadband deployment varies across years.

regulation across states. For example, the electoral college system encourages presidential candidates to pay special attention to some states (Lizzeri and Persico, 2001; Strömberg 2008). Similarly, the incentive to donate to a Congressional candidate is a function of the electoral stakes of the election she is participating in (Cox and Munger, 1989). State-time dummies can also account for the technological improvements that have helped significantly drive down the cost of providing broadband service over time (Xiao and Orazem 2011). Lastly, these variables capture the influence of state-level regulation on the construction of broadband infrastructure (Faulhaber 2002; Wallsten 2005).

The time-varying area-level characteristics $\mathbf{X}_{it}$ control for the demand and supply factors suspected to affect broadband deployment: median age, log(density), log(population), log(farm size), log(value farms), log(income), log(value housing), log(federal expenditures), % high school grads, % urban, % white, % black, % hispanic, % income $150k+, % income below poverty line, % 18+ population, % 65+ population, % male.[32] This choice of controls reflects the existing state of knowledge of the literature on the drivers of broadband demand and supply. Higher income, more educated, and younger individuals have been documented to be more frequent PC users (e.g. Jimeniz and Greenstein 1998, Goolsbee and Klenow 2002, Greenstein and Prince 2006). There is also some evidence of a digital divide disfavoring females and members of minorities (e.g. Flamm 2005). Finally, broadband service is cheaper to provide to dense urban areas, given the significant fixed costs of laying out cable infrastructure (e.g. Greenstein 2004).[33]

The specification also accounts for the possibility that a correlation between the instruments and broadband growth reflects the influence of other demographics. As shown in Section 1.3.2, broadband was not introduced at the same pace in different counties. To account for this potential confound, I include in $\mathbf{X}_{it}$ interactions between time dummies and pre-broadband values of the key demographics identified in Section 1.3.2: median age, income, federal expenditures, population, log

[32]Zip code and county level data on demographic controls for 1980-2008 comes from the IRS and US Census. Note that for each series data are unavailable for selected years of the sample. Data on median income is available every year between 1995 and 2008 (except 1996), while other data series are available in Census years (i.e. 1980, 1990, 2000, 2010). At the zip code level, population data is available for Census years (1990, 2000, 2010), and IRS data is publicly available for three years (1998, 2001, 2008). To fill these gaps, I interpolate or extrapolate. The IRS data is only available for zip code areas that house enough individuals so that reporting summary statistics is unlikely to infringe on confidentiality. Extrapolation has the disadvantage of sometimes producing out-of-range values (e.g. income below $0, % high school above 1). In these individual cases, I either exclude the particular data point from the analysis, or cap it at the maximum. For instance, negative values for zip code level income are set as missing, while a percentage of high-schoolers greater than 100% is replaced with 100%.

[33]There may, however, be greater demand for broadband in (isolated) rural areas, as the Internet can help them find additional retail options or introduce them to new education possibilities (Hindman, 2000, Sinai and Waldfogel, 2004).

farm value, % black, % below poverty line, % high school grads, % income $100k+, % urban, and % pop 18+ and % pop 65+.[34] I also include in X similarly-constructed time dummies for the dependent variable Y (e.g. campaign donations). The addition of these controls allows areas with different baseline characteristics to follow different paths in broadband availability.

1.4 Did broadband availability affect U.S. politics?

1.4.1 Preliminary analysis

I start by examining trends in the raw data on county-level presidential votes and broadband availability (Figure 1.3). In Panel A, I consider the two subsamples of U.S. counties with the highest ("top 10 percent") and lowest ("bottom 10 percent") broadband availability in 2000. The difference in broadband availability in these two sets of counties is comparable to the average broadband deployment across all counties in 2008.[35] For this reason, an inspection of the relative trends in political participation in "top 10 percent" vs. "bottom 10 percent" counties is informative, both qualitatively and quantitatively, regarding the magnitude of the effect of broadband growth on political participation. In Panel B, I divide the full sample of U.S. counties into two groups, depending on whether their broadband availability in 2000 was in the "top 50 percent" vs. "bottom 50 percent".

Because Figure 1.3 can provide both a qualitative and quantitative estimate of the effect of broadband, I focus on interpreting it in what follows (the same qualitative patterns hold in Figure 1.3). From 1980 to 1996, turnout in "bottom 10 percent" counties was 8-10 points higher than it was in "top 10 percent Internet" counties, while the share of Democratic votes was on average 3 points lower. Starting in 2000, voter turnout increased in "top 10 percent" counties but remained constant in "bottom 10 percent" counties; by 2008, voter turnout was 2 percentage points higher in the "top 10 percent Internet" counties. The same general pattern can be observed for the share of Democratic votes; by 2008, Democratic vote share was 17 percentage points higher in "top 10 percent" counties. From a visual inspection of these figures, we can thus conclude that there was

[34] All the main findings are robust to including interactions with the full list of controls, though estimates become a bit less precise.

[35] In 2008, the "high Internet" counties housed average of 11.5 Internet Service Providers. In contrast, the "low Internet" counties housed an average of 5 Internet Service Providers. The difference in broadband availability is roughly equivalent to the 7.98 growth in broadband availability in the full set of counties between 1996 and 2008.

sharp breaks in voting patterns after broadband was introduced in 2000. Furthermore, counties with and without broadband available are different in their pre-broadband turnout, which suggests that they differ on observable characteristics.

It is also instructive to examine first-stage and reduced-form regressions. In an OLS framework, this can provide evidence that the claimed relationship between topography and broadband availability / political outcomes is a robust feature of the data. In Table 1.4, I report estimates from the full county-level specifications of Section 1.3.3, which include county and state-time dummies, county controls, as well as county time trends. In an additional empirical exercise, I investigate whether the effect of topography on broadband availability varies with time. This provides a formal empirical investigation of the effect of topography on the final level vs. pace of broadband deployment (see Section 1.3.1).

I find a highly statistically significant positive correlation between land elevation and the main variables of interest. A county with a one standard deviation higher land elevation has on average an 0.18 additional ISP in a county, a 0.38 percentage-point higher turnout, and a 0.30 percentage point higher Democratic vote share. There is also a negative relationship between flow accumulation and the variables of interest. However, this negative correlation is only statistically different from zero when the dependent variable is the number of ISPs; a one standard deviation higher flow accumulation is associated with an additional 0.05 ISP. Because of this weak relationship, I do not include this potential instrument in the main analysis (see footnote 30).

Further investigation reveals that the effect of topography on broadband availability differs across years. Relative to 1996, a one standard deviation higher land elevation predicts an additional 0.05 ISP by 2000, an additional 0.25 ISP by 2004, and an additional 0.19 ISP by 2008. Hence the comparative broadband advantage provided by topography peaks in 2004, but is somewhat reversed by 2008. This suggests that favorable topography propelled broadband deployment in a certain group of counties both in the short-term and long-term.

1.4.2 Main regressions

The analysis of presidential elections is carried out at the county level. Throughout, I limit my attention to the four presidential elections between 1996 and 2008. The analysis of campaign donations is carried out at the zip code level for both presidential and mid-term election cycles between 1994 and 2008. For 1996 (and earlier years), I can confidently impute null broadband availability for all zip codes, since high-speed Internet had yet to be commercialized. Data is excluded for 1998 because of the lack of availability of broadband data (see robustness analysis in Section 1.4.4).

Table 1.5 presents estimates of the effect of broadband availability on county-level measures of voter turnout, the share of votes cast for the Democratic nominee, and the share of votes cast for the Republican nominee. Table 1.6 presents estimates of the effect of broadband availability on zip code-level aggregates of the dollar amount in campaign contributions. The first group of regressions includes all election cycles between 1994 and 2008 (excluding 2008). The second group of regressions splits donations into presidential (1996, 2000, 2004, 2008) and non-presidential (1994, 2002, 2006) election years. These empirical exercises justify the inclusion of both a mid-term and presidential election cycle (1994 and 1996).

In all regressions, I include geographic area (i.e. county or zip code), and state-time fixed-effects. For each political outcome, I report regressions from this basic fixed-effects specification. I then turn to models that also include land elevation as an instrumental variable. Time-varying demographics and interactions are as mentioned in Section 1.3.3. In models of voter turnout, I also control for the absolute difference between Republican and Democratic vote share ("winning margin"). This last variable measures the competitiveness of a race, and has been shown to have a strong effect on turnout (e.g. Strömberg 2004, Gentzkow 2006).[36] In models of campaign contributions, I also control for zip code level tax and population data.[37] Furthermore, the larger cross-section of zip codes (relative to counties) makes it possible to test whether the effect of broadband changes with time.

In the most basic specification, each new ISP brings a 0.53 percentage points increase in voter turnout, a 0.66 percentage points increase in Democratic vote share, and a $12,287 increase in campaign donations. When the model is re-estimated treating ISPs as endogenous, the effect increases relative to the fixed effects estimates to 0.87 points for voter turnout (64 percent increase), 1.48 points for Democratic share (124 percent increase), and $34,378 for campaign donations (180

[36]Note that I do not include winning margin in models of Democratic vote share. This is because of how closely related its definition is to the dependent variable (Democratic vote share). In any case, results are unaffected by the including of voting margin as a control.

[37]I also include controls for missing values of the salaries and returns variables.

percent increase). Breaking down the effect of broadband on campaign donations across election cycles, I find that its effect is three times larger in presidential election years ($46,977 vs. $17,618).[38]

With the addition of county controls, the broadband estimate changes relative to the specification with fixed effects and the instrumental variable. The effect of broadband availability now becomes 2.15 points on voter turnout (148 percent increase), 1.62 point on Democratic vote share (9 percent increase), $7,333 on donations in non-presidential election cycles (58 percent decrease), and - $251.4 on donations in presidential election cycles.[39] Furthermore, the effect of an additional ISP on campaign contributions is larger between 1994 and 2002 than it is between 2002 and 2006. This suggests the effect of broadband availability declines over the sample period.

One implication of these estimates is that broadband benefitted both Democrats and third-party candidates, to the detriment of Republicans. The effect on third-party candidates is consistent with Campante et al.'s (2013) study of Italian elections, who find that broadband Internet benefitted the newly formed M5S party. Another implication is that omitted variables artificially reduced the size of the broadband estimate for presidential election outcomes. To investigate the drivers of these effects, I examine coefficient estimates on the control variables in the specification (not reported). First, in the 2000s racial minorities received lower broadband access, but their turnout increased at a faster pace than the rest of the population. Second, older individuals have shifted their support more drastically toward Democrats, while receiving lower broadband demand. Third and last, the increase in turnout and Democratic support was less pronounced in dense urban areas, which concurrently have had more broadband available.

The estimate of a heterogeneous effect of broadband on donations in presidential vs. non-presidential election cycles motivates further investigation. I re-estimate the most conservative regression of Table 1.6 separately for different groups of donors / recipients of campaign money. In Table 1.7, I divide donations depending on recipient characteristics, for both presidential and non-presidential election years. In Table 1.8, I divide donations on the basis of donations / donor characteristics, focusing on non-presidential election years. To avoid cluttering the tables, I only report the coefficient estimate on broadband availability. Notably, I also draw a distinction between

[38] Residents in the average zip code area donated 50 percent more money in the 1996 presidential election year than they donated in the 1994 presidential election cycle. This general pattern is present across election years. This makes sense because the presence of a presidential race implies a greater number of political campaigns to give to.

[39] Consistent with further reduction in attenuation bias, when adding flow accumulation as an instrument, estimates are more precisely estimated, and slightly higher.

donations going toward House candidates running in the same state as the donor, and donations going toward House candidates running outside the state of the donor.[40] A comparison of in-state vs. out-of-state House donations provides a direct test for whether the salience of political races interacts with the effect of broadband. In fact, in the past information about House races was only accessible to individuals living in the same state, as this type of political activity was only covered by local newspapers (e.g. Snyder and Strömberg 2010).

In non-presidential elections, I find that broadband's positive effect was mainly driven by an increase in donations toward PACs and in-state House candidates.[41] As for the lack of broadband effect in presidential election years, evidence suggests that it masks a large re-allocation of funds from in-state House candidates and PACs toward presidential candidates (particularly Democrats). I also find that broadband availability did not affect the amount donated to House candidates running out-of-state. This provides an interesting contrast with recent survey evidence that consumers of online news are more likely to be interested in niche political issues (Nie et al. 2010).

The results for campaign contributions for particular demographic groups show varied effects of ISPs expansion on the amount contributed. The broadband effect is larger on female vs. male contributions (in percentage terms), even though evidence suggests the existence of a gender gap in Internet use that favors males (NES, 1998; Bimber 2000). Moreover, the broadband effect is concentrated on smaller donations ($\leq$ \$1000). Keeping in mind that donors who give less than \$200 are excluded from the dataset, this bolsters, and precisely quantifies, received wisdoms about the stimulating effect of the Internet on small donations. The model results suggest that an additional ISP in a zip code area causes a large increase in contributions in the donations of professionals. As the elderly are infrequent Internet users, one would expect to find a smaller effect of broadband on their contributions (if the specification is valid). Reassuringly, I find that broadband leads to no changes in the contributions of these retired individuals. But turning to further robustness analysis, I interpret the magnitudes of my findings in the next subsection.

[40] The phenomenon of donating out-of-state has recently attracted attention because of concerns of undue influence in the electoral process, as well as their increasing prominence. In 1995, Senator Feingold attempted – but failed – to introduce legislation to curtail the scope of donating out-of-state. In 2010, the practice of donating out-of-state attracted media attention when Representative Dennis Kucinich raised 98.2 percent of his campaign funds from outside of his state. Only a few potential determinants have been suggested for this type of donations (influence over the legislative process (Grenzke 1988); competitiveness of the race (Gimpel et al. 2008); interconnectedness of the donor's geographic area (Bednar and Gerber 2012).

[41] I exclude Senate / gubernatorial elections as they involve seats open for elections once every six years, which substantially complicates the implementation of a fixed effects framework.

1.4.3 Interpreting magnitudes

The above results provide evidence for whether political behavior changes in response to access to broadband Internet service. To interpret the magnitude of the findings, it is useful to put them in relation with contemporaneous forces affecting voting / donating, and to compare them with related estimates from the literature. I compute persuasion rates on voting behavior, i.e. the share of individuals affected by broadband Internet, among broadband users. Since donating to political campaigns is a much less frequent political act, I instead compare broadband's effect with that of contemporaneous forces that have also affected campaign donations.[42]

To assess the importance of broadband availability on voting behavior, I compute its persuasive effect for three potential messages: "Get out the vote" (GOTV), "Vote for the Democratic candidate" (Positive message), "Do not vote for the Republican candidate" (Negative message). To compute these statistics, one could assume a uniform 63.5 percent exposure to broadband Internet for all subpopulations.[43] However, this approach ignores two crucial considerations, which may lead to an overestimate of persuasion rates. First, broadband access was higher for non-voters than voters, which would lead to the overestimation of the effectiveness of the GOTV message.[44] Second, anecdotal evidence suggests that online political campaigns have relied heavily on the convincing effort of volunteers within their social circles. In the words of John Kerry's director of Internet strategy Josh Ross, "What we were trying to do is have a multiplier effect. If you can recruit one volunteer, that volunteer can contact many, many voters as they show up for their local canvas event" (Vaccari 2008). In order to avoid overestimating the effect of broadband, a more cautious approach would be to assume 100 percent broadband access. This implies persuasion rates of 29.5 percent, 9.63 percent, and 12.19 percent for the GOTV, positive, and negative messages, which

[42] The data show that the average zip code houses an average of 21 donors who each donated about $1,000. Donating thus only involves a very select group of individuals (0.3 percent of the adult US population). In contrast, in the average county voting about half of the adult population voted in 1996. This indicates that the representative donor and voter are two very different individuals.

[43] The estimate comes from October 2009 Census data. See government's report as http://www.ntia.doc.gov/files/ntia/publications/ntia_internet_use_report_feb2010.pdf (retrieved on October 23, 2013).

[44] Assuming uniform broadband access potentially leads to an overestimate of the persuasiveness of broadband, particularly its GOTV message. First, non-voters exhibit higher levels of broadband access than non-voters. Dense urban areas have more broadband access and house significantly more non-voters (see Figure 1.3). Furthermore, the youth have consistently lower turnout than other segments of society, as well as greater Internet use. The youth constitute a large fraction of the electorate in any given election. Using age-specific data on mortality and population growth from the Human Mortality Database, I estimate that 31.8 percent of the 2008 voting age population consists of individuals who were too young to vote in 1996.

should be interpreted as lower bounds (respectively).[45]

The relative magnitudes of these three effects suggest that broadband's GOTV message was more persuasive than its partisan messages. In comparison to the rest of the literature, the effect of broadband Internet on U.S. voter turnout is somewhat larger. In their survey of the literature, DellaVigna and Gentzkow (2010) report that persuasion estimates on voter turnout range between 4.3 percent and 20.5 percent. This suggests that broadband Internet was particularly effective in encouraging non-voters to cast a vote. There are only a handful of other studies that have established an effect of the media on vote shares (DellaVigna and Kaplan 2007; Gerber et al. 2009, Enikolopov et al. 2011). Relative to these studies, broadband Internet's positive and negative messages are intermediate (range for positive messages: 8 percent to 20 percent; range for negative messages: 14 percent to 65.4 percent).[46]

To assess the importance of broadband availability for campaign donations, I compute the share of the increase in campaign donations due to broadband. To do so, I rely on estimates from the most conservative specifications of Table 1.8. Overall, I estimate that broadband increased donations by an average of 144 percent in non-presidential election years, and decreased them by 5 percent in presidential election years. Over the time period of the study, the total increase in donations was 147 percent in non-presidential election years, and an astonishing 278 percent in presidential election years. Taking a dollar-weighted average, I then find that broadband explains 41.5 percent of the increase in individual contributions over the time period of the study.

At least part of the remaining variation is likely to have come from the 81 percent increase in GDP over the time period. In fact, the evolution of campaign donations has historically mirrored

[45] An additional ISP is associated with a 2.15 percentage points increase in voter turnout, a 1.62 percentage points increase in Democratic vote Share, and a 1.95 percentage points decrease in Republican vote share (Table 1.5). Note that in 2008 average number of ISPs in a county was 7.98, average turnout was 58.83 percent, average Democratic share was 41.49 percent, and average Republican share was 56.86 percent. I compute persuasion rates adapting formulas derived by Enikolopov et al. (2011): $f_{\text{GOTV}} = \frac{\widehat{\beta}_{\text{turnout}} * ISP_{2008}}{1-(t_{2008}-\widehat{\beta}_{\text{turnout}} * ISP_{2008})}$, $f_{Positive} = \frac{\widehat{\beta}_{\text{Democrat}} * ISP_{2008} * t_{2008}}{1-\left(d_{2008}-\widehat{\beta}_{\text{Democrat}} * ISP_{2008}\right)}$, and $f_{\text{Negative}} = \frac{-\widehat{\beta}_{\text{Republican}} * ISP_{2008} * t_{2008}}{r_{2008}-\widehat{\beta}_{\text{Republican}} * ISP_{2008}}$.

[46] The study that finds the largest negative effect of a media was authored by Enikolopov et al.. They study the effect of the introduction of an independent news television channel, NTV, on Russian politics. They find a large negative effect of NTV on Putin supporters' voting behavior. A larger effect in Russia provides strong support for the theoretical prediction that government capture of the media becomes less likely in a competitive news market (e.g. Besley and Prat 2006). In Russia, the media has historically been government-controlled, which may give a single independently-owned media outlet the power to reveal new information to society. That being said, even in countries with strong constitutional guarantees for press independence, the state has many levers by which to influence it. For example, less than a decade ago (in 2004), the Joint Chiefs of Staff successfully pressured a CBS reporter to delay reporting on the Abu Ghraib prison scandal.

GDP growth (Ansolabahere at. al 2003).[47] However, the discussion thus far still leaves the increase in campaign donations in presidential election cycles largely unexplained. I hypothesize that the 2002 Bipartisan Campaign Reform Act (BCRA) may have played a role. The BCRA discouraged spending on unregulated electioneering communications, and raised contribution limits on regulated donations (see Table A1 for the evolution contribution limits). In general, PACs are the institutions that carry out political advertising. One way to provide support for the claim that the BCRA helped fuel the increase in campaign donations is thus to examine the evolution in PACs donations. Consistent with this interpretation, 50.2 percent of the massive increase in presidential years came from a surge in donations toward PACs.

1.4.4 Robustness analysis

I note three concerns that are specific to the campaign donations analysis, but do not apply to the presidential voting analysis. The first concern comes from the exclusion of the 1998 election cycle, a year of transition where broadband might have been introduced. This choice was motivated by the lack of broadband availability data for that year. If an increase in campaign contributions was already underway by then, then I may be misattributing a secular time trend to broadband in the main analysis. To alleviate this potential concern with the specification, I instead assume some value of broadband for 1998, e.g. zero or half of the 2000 value. I report estimates in Table 1.10; regardless of the specification used, the broadband effect persists and is similar in magnitude to the earlier results.

The next two concerns both stem from the existence of zip codes whose residents contribute zero dollars to political campaigns. Table 1.9 divides the universe of 27,405 zip codes into 8 bins labelled by a number between "0" and "7", which corresponds to the number of election cycles for which at least one zip code resident donated money.[48] There are 1,488 zip code areas in which no person ever contributed over the sample period. Furthermore, half of all zip codes donate no money to politics in at least one election.[49]

[47]This pattern has held robustly until the 1990s, and has been interpreted as evidence that donations constitute a normal consumption good.

[48]Naturally, zip code areas with larger populations contribute more often to politics. Moreover, a striking 95 percent of all donations come from 14,622 zip codes which donate in each election cycle between 1994 and 2008.

[49]The sample would suffer from severe selection bias if I dropped all data points with zero campaign donations and focused on a balanced panel of observations. In fact, included observations would by design have a greater propensity for giving.

This large presence of zeros support the use of the dollar amount of donations – rather than its logarithm – as the dependent variable in the main analysis above.[50] But one potential concern with the dollar variable is its non-normal distribution. As evidence against this claim, Appendix Figure A1 suggests the existence of no such systematic relationship between the size of the residuals and fitted values.

This large presence of zeros may also reflect a systematic omission by the FEC, as it is surprising that some large zip codes never contribute to politics. To make this concern concrete, there are 257 zip codes with at least 1,000 inhabitants with zero donors. If one assumes that no such such zip code area should exist, this indicates that the FEC data is somewhat incomplete. In Table 1.10, I find that the effect of broadband is both qualitatively and quantitatively robust to the exclusion of the 1,488 problematic zip codes.

1.4.5 Does the effect of broadband mask an effect of cable television?

One may be worried that the analysis misattributes to broadband an effect of cable television.[51] As previously mentioned, the broadband market has been dominated by cable television firms. Those firms had an early advantage in offering broadband service, since they only needed to upgrade part of networks to make them broadband compatible (see Section 1.3.1 and the Appendix). The deployment of broadband infrastructure is thus likely to have exhibited similarities with the earlier deployment of cable television. This would be a source of concern if cable television's effect on political participation was similar to the one attributed above to broadband.

To test the hypothesis that cable television confounds the effect of broadband, I explore whether geographic variation in broadband availability predicts variation in political participation *before* the advent of broadband. I use broadband availability as a proxy given the lack of geographically disaggregated data on cable television penetration.

The core analysis in this subsection uses data on county-level presidential elections outcomes,

[50] Using a logarithm specification is an unsatisfactory approach since requires making an ad hod assumption about the zeros in the data. I experiment with such specifications (e.g. adding one cent, one dollar, 100 dollars), but I omit reporting these findings because of their sensitivity of my findings to the assumption made.

[51] Note that the deployment of dial-up followed different patterns from broadband, which makes it an unlikely confounder (e.g. Grubesic 2006). Furthermore, a lack of effect of dial-up is consistent with anecdotal evidence on campaign donations. In fact, the first online fundraising campaigns coincided with the time of the introduction of broadband. were led by Senator McCain in 2000 ($2 million) and Howard Dean in 2004 ($41 million). In 2008, only 37 percent of dial-up users consumed online political news, relative to 62 percent for broadband users (Tedesco 2011).

whose time frame extends back to 1980. This long panel allows me to analyze outcomes in two separate pre-broadband periods that are of similar length to the one used in the main analysis: 1984-1996 and 1980-1992. Because the deployment of traditional cable networks took place in the 1980s, this time frame is appropriate for an investigation of the effect of cable television on politics. In fact, the percentage of household with cable access increased from 23 percent to 63 percent between 1980 and 1996 (see Figure 1.5).[52] Therefore, one would expect any effect of cable television to have been more pronounced in these years than in the 2000s.

In Table 1.11, I report regression estimates on these data using the most conservative empirical specification from Table 1.5 (i.e. with full controls and the instrumental variable). I find that broadband availability does not predict a significant change in Democratic vote share between 1980 and 1984-1992, or voter turnout between 1984 and 1988-1996. However, I find that the growth in the number of ISPs in 1996-2008 predicts a *decrease* in voter turnout between 1980 and 1984-1992, as well as a *decrease* in Democratic share between 1984 and 1988-1996. The persistence of a negative effect into the broadband years would imply that I underestimate the effect of broadband over the 1996-2008 sample period. These findings thus only reinforce the interpretation of the main estimates of Section 1.4.2. They are also consistent with empirical evidence for a depressive effect of (network) television on voter turnout (Gentzkow 2006). More broadly, they conform with the general decrease in turnout observed between 1960 and the end of the twentieth century (see Figure 1.4).

The secondary analysis uses data on zip code-level campaign donations. The time frame for these data only extends back to 1990. I estimate a model of the change in campaign donations between 1990 and 1998 on the growth in broadband availability between 1994 and 2002. In Table 1.11, I report estimates using the specification from the last column of Table 1.6. I find that broadband does not significantly predict the evolution of pre-broadband donations. If we assume that cable television indeed drives these findings, we may deduce that some its influence had faded away by the early 1990s. This would be consistent with the fact that cable television penetration had almost peaked by that time. Extrapolating this finding into the future, we would expect any cable

[52] The increase in cable penetration over this time period must have at least been partly driven by the large expansion in the number of channels. While in the early 1980s, only a handful of programming networks existed, programming options increased drastically to more than 150 channels by the mid 1990s (e.g.: CNN, USA, MTV); see http://www.ncta.com/About/About/HistoryofCableTelevision.aspx (retrieved on October 20, 2013).

television effect to disappear by the 2000s, as cable television penetration continuously decreased from 1997 onwards (see Figure 1.5).

1.5 Through what channel(s) did broadband availability affect U.S. politics?

1.5.1 Methodology

As a final analysis, I explore the mechanism(s) behind the effect of broadband availability on U.S. politics using individual-level survey data. I use survey data that measure how much people know about politics, their moral values, the medium for campaign donations (i.e. online vs. offline), and the medium for political news consumption (e.g. television, radio). These data all come from the Pew research center, which regularly fields surveys on nationally representative samples of individuals. The Appendix contains details about data construction, as well as a full list of the survey questions used as dependent variables in the forthcoming analysis.

Over the course of the 2007-2008 presidential election campaign, Pew fielded five surveys to assess the political knowledge of respondents. To do so, they asked questions about issues that either pertain to general political knowledge, or have recently been featured in the news. For example, one question asks individuals to name the vice-president of the United States, while another asks them to estimate the number of soldiers who died in Iraq. I construct an index of political knowledge based on the share of correct answers provided, giving equal weight to each survey question.[53]

The analysis of media consumption is based on ten surveys spanning three presidential election years (2000, 2004, 2008). The main question asks individuals to list their preferred source for obtaining news about election campaigns (e.g. Television, Internet). A follow-up question, only available in 2004 and 2008 surveys, asks respondents who listed "Television" as a preferred media to indicate the television channel they most often turn to. Using answers to these two questions, I construct media-specific measures of news consumption, set to "100" when the media in question was mentioned by the respondent, and "0" otherwise.[54]

[53]Formally, this index is defined as knowledge index$_{i,s} = \frac{\#\text{ correct answers}_{i,s}}{\#\text{ questions}_s}$ for respondent i participating in survey s.

[54]Respondents can volunteer multiple answers to each question and so news consumption does not in general add up to "100" across media. I am reluctant to give an intensity interpretation based on the order in which media are mentioned, or the total

Finally, the analysis of liberal values is based on sixteen Pew surveys fielded between 2003 and 2008, which ask individuals if they would support the legalization of abortion or gay marriage. The question on gay marriage is available in fifteen surveys, while the question on abortion appears on only the six most recent surveys (fielded between early 2007 and late 2008). For the five surveys that ask the two questions, I construct a "liberal index", defined as the share of answers provided that are pro-gay or pro-choice.

To estimate the effect of broadband availability, I match each respondent with the number of ISPs in her zip code or county of residence. I control for possible confounding factors with the following list of demographics, along with a corresponding set of dummies indicating missing values: age, party affiliation, employment status, gender, urban / suburban / rural area, income, race, educational attainment, home ownership, marital status, life satisfaction, household size, household head, religion. However, I abandon both the instrumental variables and fixed effects framework of the last section.

The different empirical specification for this Section is motivated by the fact that I neither have data for pre-broadband years, nor repeated observations for the same individuals. I do not include zip code / county dummies because Pew samples are only meant to be *nationally* representative, making it conceptually inappropriate to construct zip code or county aggregates. As for the instrumental variable strategy, it was only designed to identify the effect of broadband relative to pre-broadband years, and the Pew data starts in 2000.

Attenuation bias through measurement error is a more important concern than in the main analysis. In fact, the analysis matches a zip code / county measure of broadband availability to individual respondents who do not in general represent the behavior of the average resident in a geographic area. I thus relegate to a robustness analysis the addition of state-time dummies and zip code / county- level controls (Table A4). While attenuation bias reduces the effect of broadband in a few cases, the main findings are generally robust to the inclusion of these controls.

---

number of media listed. In fact, someone who only mentions Media A may consume less Media A-news than a person who lists both Media and Media B. For instance, providing multiple answer choices may signal stronger interest in politics, and hence greater total media consumption.

1.5.2 Results

In Tables 12-14, I report regression estimates. I interpret the difference in the behavior of respondents living in a zip code / county with a one standard deviation higher number of ISPs. I find that a one standard deviation higher broadband availability predicts a 3.5 percent higher political knowledge, a 9.9 percent / 4.9 percent more likely pro-gay / pro-choice attitude, and a 30.6 percent higher chance of donating online. All estimates – except the one on pro-gay attitude – are solely based on data from the 2007-2008 election cycle. As argued above, the number of ISPs is measured with error in this setting. To the extent that I am dealing with classical measurement error, I would be understating the full extent of the effect of broadband.

I then find that a one standard deviation higher number of ISPs predicts a 41 percent higher Internet news consumption, and a 3.6 percent lower television news consumption. However, broadband has no detectable effect on other forms of news acquisition (radio, magazines, radio). Hence greater broadband availability crowds out television news, and increases the consumption of Internet news. It should be emphasized that a positive effect of broadband availability on online political news consumption was not self-evident a priori. The Internet could in principle have *reduced* the intake of political information. For example, Falck et al. (2013) provide evidence in the context of Germany that the introduction of broadband Internet has reduced interest in politics, by creating new forms of entertainment.

The next empirical exercise examines whether the promotion of liberal values was the result of online media bias, or online information acquisition. To do so, the broadband effect is interacted with the ideology of the respondent. If the Internet had a liberal bias, a rational Republican-leaning individual should react by not consuming online news. As evidence against either rationality of liberal bias of the online media, I find that Republicans were in fact *more* likely than Democrats to consume online news (see Gentzkow and Shapiro 2011 for consistent evidence). This suggests the importance of reconsidering the effect of biased news on viewers.

Finally, I ask which TV channels were crowded out by online political news. The decrease in interest in television news affected both cable television and network / local television, and in particular FOX news and CBS news. These findings provide additional evidence for the claim made in Section 1.4.5 that the number of ISPs does not capture pre-existing trends in cable television deployment.

1.6 Conclusion

In this chapter, I have studied the effect of broadband Internet on political participation. I found that the Internet is slowly emerging as an alternative to television for the retrieval of political news. Moreover, I documented large effects of broadband availability on campaign donations, voter turnout, and support for the Democratic party. As plausible mechanisms for these findings, I then showed that broadband availability was linked with an increase in online political news consumption, online donations, political knowledge, and liberal values (see Figure 1.6).

The analysis is, however, far from exhaustive, and it suggests a number of fruitful areas for further inquiry. First, it is important to understand the effect of broadband on voting in Congressional elections. In fact, evidence suggests that extrapolating conclusions about the effect of a media across political contexts may lead to flawed conclusions (e.g. Gentzkow 2006). Consistent with this interpretation, I have provided evidence for a great deal of heterogeneity for the effect of broadband on campaign donations across political causes. In order to make progress in this direction, it will be necessary to collect county-level data on Congressional elections for the 1990s, as it is not readily available for this time period.

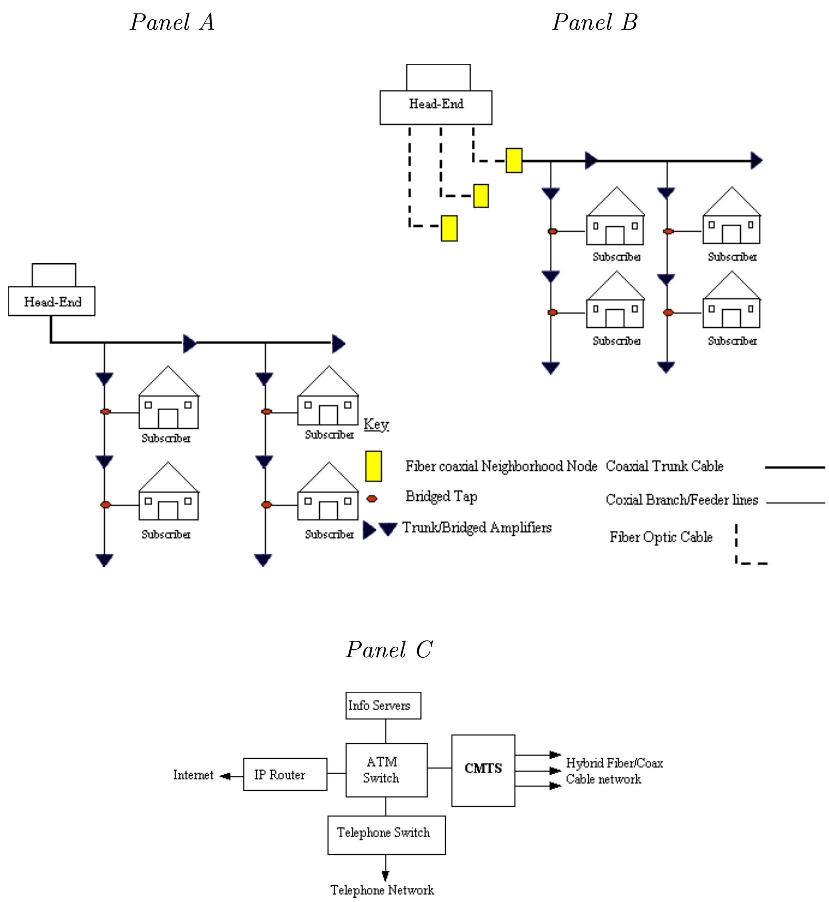
The documented effect of broadband also points toward the importance of incorporating behavioral mechanisms for persuasion in the political economy literature (e.g. DeMarzo et al. 2003, Eyster and Rabin 2009). In particular, Republican audiences' heavy reliance on online news is hard to reconcile with rational behavior, as individuals should be able to filter out the bias of media with a known partisan inclination, e.g. by switching to alternative news sources (e.g. Gerber at. al 2009, Chiang and Knight 2011, Durante and Knight 2012). In order to generate more broadly applicable conclusions for the effect of media bias, one possibility would be to treat differently subtle (i.e. online) and explicit (i.e. television) bias in political news.

The findings in this chapter can also help inform, and relate, policy debates on campaign finance regulation and broadband open-access rules. The low broadband connectivity of disadvantaged populations and the increasingly large role of corporate donors have often been deplored (e.g. Lessig 1999, 2011). However, the intrinsic connection that exists between broadband access and campaign donating has not been formally recognized in the past. My study suggests that broadband access can help tilt the balance of power in favor of individual donors. This finding is timely, given the renewed concerns about the undue influence of corporations and the wealthy following *Citizens United vs. FEC* (2010).

More broadly, this study provides a portable framework for studying the effect of the Internet. The Internet is gaining an increasingly central status in all wakes of social and political life. Furthermore, its effect is likely to evolve with time, as broadband access spreads more widely, and the quality of service improves. However, the lack of clear theoretical predictions have opened the gate for endless hypothesizing and generalizing over its effect. The Internet may have been a bane, or a boon, but the answer is farther than a Facebook status away.

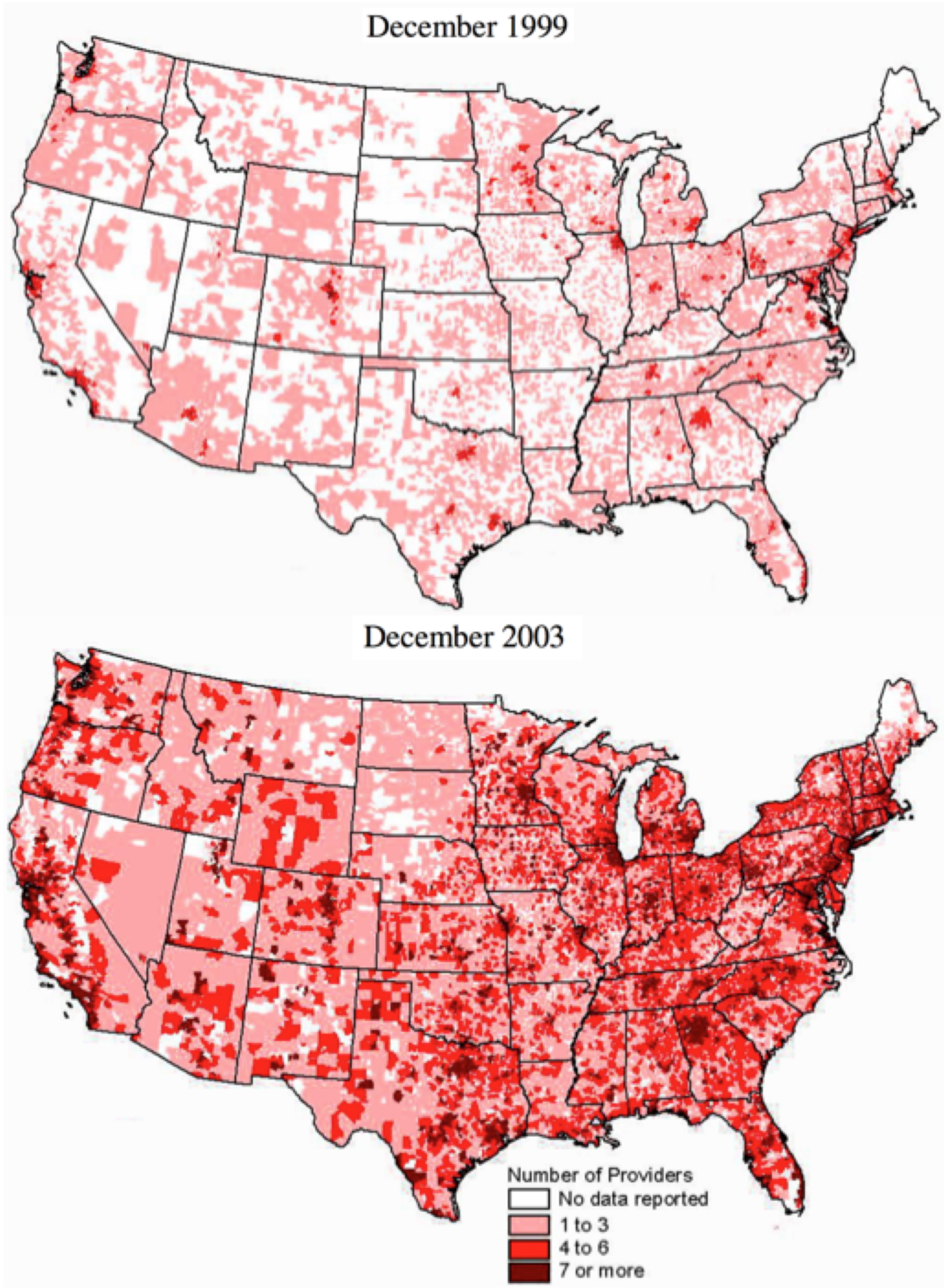
1.7 Figures

Figure 1.1: The evolution of cable television networks



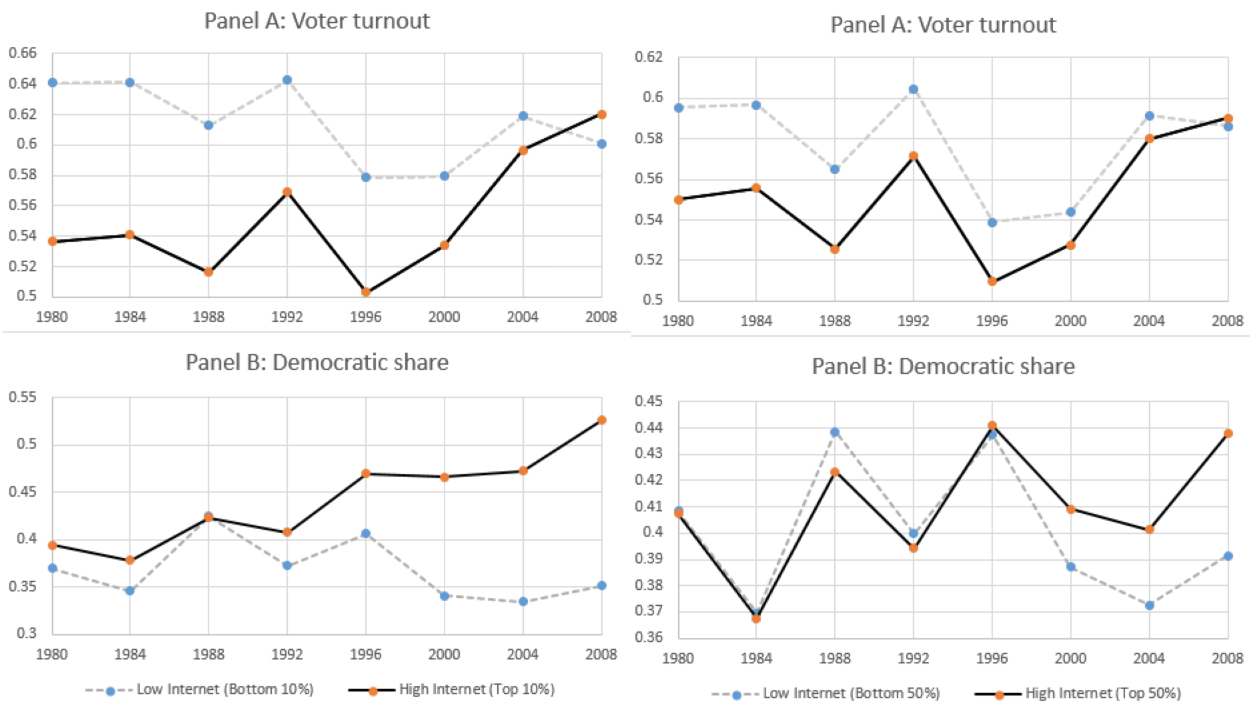
Notes: Panel A: A traditional cable network consisted of a Head-End connected to a trunk cable (in bolded black line) that enters various sections of a city. The trunk cable was itself connected to a number of branch networks (light black lines) that connect to taps (in red), which enter the subscribers homes. Panel B: A modern cable network differs from a traditional cable network in a few ways. The Head-End is connected to multiple fiber optic cables (dotted lines), which themselves connected to fiber optic neighborhood nodes (in yellow). Panel C: The Head-End of a broadband cable television network hosts broadband equipment, which includes a Cable Modem Termination System (CMTS). Source: Vicomsoft

Figure 1.2: Number of high-speed Internet Service Providers by zip code



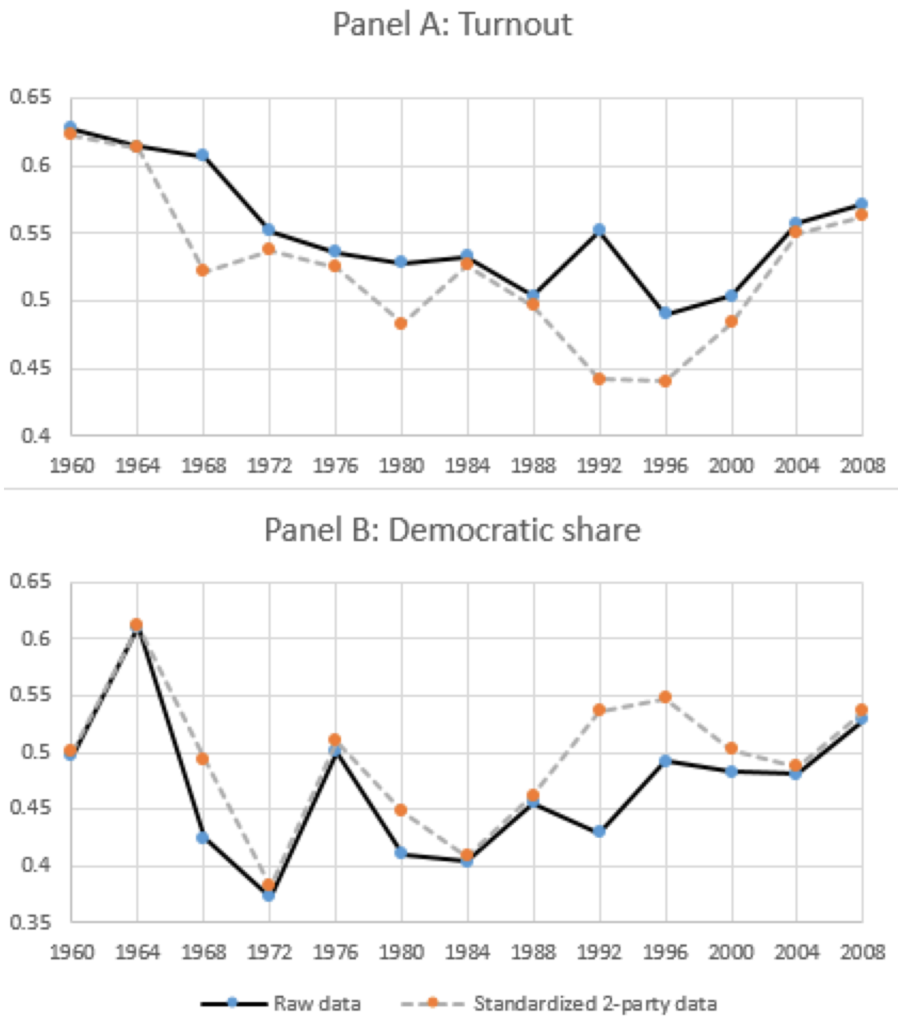
Source: Xiao and Orazem (2011)

Figure 1.3: Relative trends in voting behavior as a function of broadband access (1980-2008)



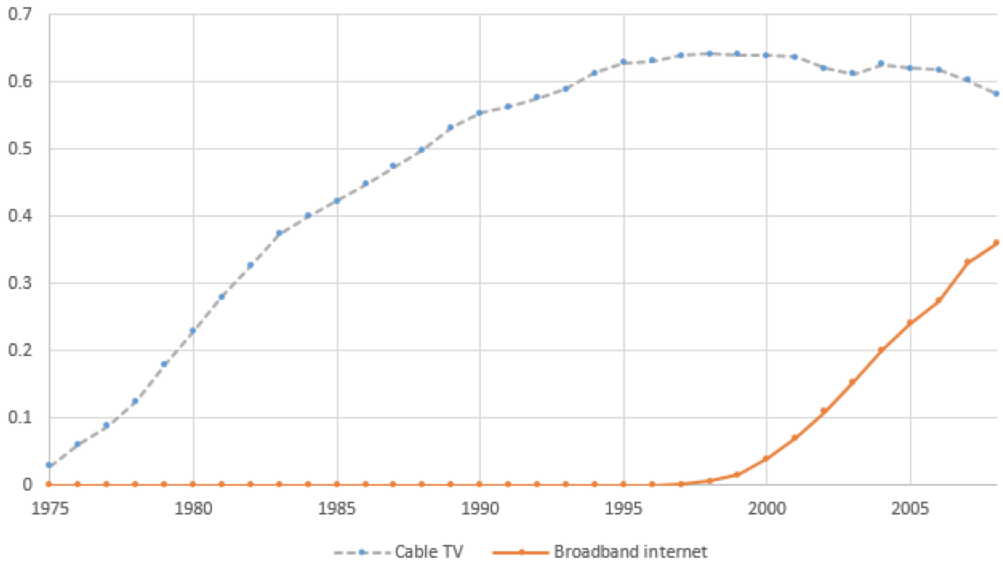
Notes: This Figure plots the evolution of Voter turnout and Democratic vote share in 1980-2008 using county-level elections data described in the text. The "High Internet (Top 10 percent)" / "High Internet (Top 50 percent)" / "Low Internet (Top 10 percent)" / "Low Internet (Top 50 percent)" data series corresponds to the evolution of political outcomes for counties whose broadband availability belongs to the highest decile / top 50 percent / lowest decile / bottom 50 percent in 2000 (respectively).

Figure 1.4: Voting behavior in presidential elections



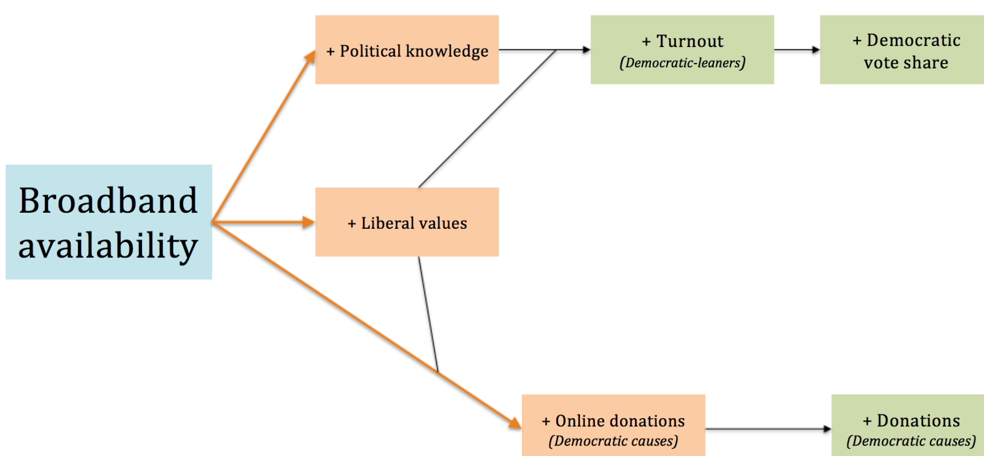
Notes: This Figure plots the evolution of Voter turnout and Democratic vote share in 1960-2008, using national level data from the FCC. The "Raw data" series count all votes, while the "Standardized 2-party" series only count votes for the two major parties (Republicans, Democrats).

Figure 1.5: Cable TV and Broadband penetration rates



Notes: This Figure plots the evolution of the percentage of American households with cable television and broadband Internet subscriptions between 1975 and 2008. Values were constructed by the author using two complementary datasets: 1) Data on the number of subscribers to cable television/broadband come from NSL Kagan (https://www.ncta.com/industry-data), and 2) data on the number of U.S. households in January of each year from the Census. Source: author's calculations.

Figure 1.6: Process for the effect of broadband Internet on political participation



1.8 Tables

Table 1.1
Relationship between broadband penetration and the number of ISPs

	% 200 kb/s +			% 720 kb/s +		
	1996, 2008			1996, 2008		
	Fixed effects	Pooled OLS	2008	Fixed effects	Pooled OLS	2008
Number of ISPs	5.42***	5.19***	2.52***	4.48***	4.38***	3.15***
	(0.0372)	(0.0385)	(0.114)	(0.0363)	(0.0374)	(0.121)
Intercept	1.10***	2.03***	25.4***	0.51***	0.93***	11.7***
	(0.149)	(0.107)	(0.957)	(0.145)	(0.0910)	(1.02)
R^2	0.865	0.765	0.127	0.825	0.710	0.174
Number of counties	3,108	3,108	3,108	3,108	3,108	3,108

Notes: Observations are at the county-level, and standard errors are clustered by county. The dependent variable is the percentage of the population with broadband access under two alternative definitions. The FCC divides counties into six groups on the basis of broadband penetration: 1) Zero broadband penetration, 2) 0<broadband≤20%, 3) 20<broadband≤40%, 4) 40%<broadband≤60%, 5) 60%<broadband≤80%, and 6) broadband>80%. For each category, I impute the middle of the range (e.g. 10% penetration for counties in the second category). For county fixed effects specifications, the intercept captures the average value of all county interecepts in the sample. *** indicates significance at the 1 percent level.

Table 1.2
Distribution of the number of ISPs

	1994, 1996	2000	2002	2004	2006	2008
10th percentile	0	0	0	2	2	2
median	0	2	2	2	5	7
90th percentile	0	4	7	10	13	12
Minimum	0	0	0	2	2	2
Mean	**0**	**1.54**	**3.04**	**4.43**	**6.13**	**7.22**
Maximum	0	10	18	20	25	23

Notes: These values come from the sample of 27,405 “standard” zip codes used in the analysis

Table 1.3

Summary statistics

	N	mean	s.d.	Mean by # ISPs in 2000 Bottom 50%	Top 50%
Main variables					
% turnout	3,105	52	10	54	51
% democrat vote	3,105	44	11	44	44
% republican vote	3,105	45	12	44	45
mean land elevation (m)	3,105	438	508	425	451
1994 campaign donations ($)	*27,405*	*13,740*	*60,365*	*827*	*22,082*
1996 campaign donations ($)	*27,405*	*21,219*	*91,350*	*1,321*	*34,073*
Control variables					
median age	3,105	37	4	37	35
population per square mile	3,105	242	1,667	41	444
county population	3,105	89,772	293,434	24,247	155,296
farm size	3,105	645	1,449	757	533
farm value ($)	3,105	2,099	4,377	1344	2854
median income ($)	3,105	36,280	8,970	32,822	39,738
new housing units value ($)	3,105	59,262	204,306	9,393	109,130
federal expenditures *(million $)*	3,105	0.48	1.7	0.13	0.83
% high school	3,105	77	8.7	75	79
% urban	3,105	40	31	26	55
% male	3,105	50	1.9	50	49
% white	3,105	85	16	86	84
% black	3,105	9	14.5	8	9
% hispanic	3,105	6	12	5	7
% high income	3,105	1	0.8	0	1
% below poverty	3,105	3	4.8	2	5
% 18+ population	3,105	75	3.2	74	74
% 65+ population	3,105	15	4.1	16	14
aggregate salaries (million $)	*27,080*	*143*	*235*	*20.3*	*220*
number of tax returns	*26,033*	*4240*	*5,771*	*808*	*6,390*
number of contributors	*27,405*	*21*	*53*	*3*	*27*
zip code population (2000)	*27,405*	*9,533*	*13,002*	*1,942*	*14,436*

Notes: Summary statistics are for 1996 unless otherwise reported. Zip code level variables are italicized, and the remaining variables are at the county-level.

Table 1.4

First-stage and Reduced form estimates

	Number of ISPs			Turnout		Dem. vote share	
	(1)	(2)	(3)	(1)	(2)	(1)	(2)
Mean elev. * $I_{t\geq 2000}$	0.137***	0.130***	0.0441***	0.295***	0.311***	0.232***	0.322***
	(0.0256)	(0.0255)	(0.0149)	(0.0960)	(0.0992)	(0.0804)	(0.0724)
Flow acc. * $I_{t\geq 2000}$		-0.0254***	-0.0102*		-0.0335		-0.0131
		(0.00981)	(0.00590)		(0.0293)		(0.0347)
Mean elev. * $I_{t\geq 2004}$			0.158***				
			(0.0339)				
Flow acc. * $I_{t\geq 2004}$			-0.0481***				
			(0.0124)				
Mean elev. * $I_{t\geq 2008}$			-0.0518				
			(0.0326)				
Flow acc. * $I_{t\geq 2008}$			0.0510***				
			-0.0146				
Mean d.v. (2008)	*7.98*	*7.98*	*7.98*	*58%*	*58%*	*41%*	*41%*
R^2	0.946	0.946	0.946	0.757	0.760	0.769	0.804
Number of counties	3,105	3,105	3,105	3,105	3,105	3,105	3,105

Notes: Standard errors are clustered by county. The analysis is always based on a panel of four presidential election cycles (1996-2008). The first raw gives the name of the dependent variable used Turnout is the proportion of the voting age population who casts a vote in presidential elections: Number of ISPs, Turnout, Democratic vote share. Democratic vote share is the share of votes in presidential elections going towards the Democratic candidate. All regressions include county dummies, state-time dummies, time-varying controls, and time itneractions. Time-varying controls are the following county-level demographics: median age, log(density), log(population), log(farm size), log(value farms), log(income), log(value housing), log(federal expenditures), % high school grads, % urban, % white, %black, % hispanic, %income \$150k+, % income below poverty line, %18+ population, %65+ population, % male. Time interactions are year dummies interacted with 1996 values of the following time-varying controls: median age, income, federal expenditures, population, log farm value, % black, % below poverty line, % high school grads, % income \$100k+, % urban, and % pop 18+ and % pop 65+. Time interactions also include year dummies interacted with the dependent variable (e,g, voter turnout). In the turnout analysis, I also include winning margin and the corresponding time interactions as controls. * indicates significance at the 10 percent level., *** indicates significance at the 1 percent level.

Table 1.5

Effect of broadband on voting in presidential elections

	Turnout			Dem			Rep
	(1)	(2)	(3)	(1)	(2)	(3)	(3)
Number of ISPs	0.528***	0.867***	2.15***	0.658***	1.48***	1.62**	-1.95**
	(0.0412)	(0.0771)	(0.800)	(0.0367)	(0.0734)	(0.647)	(0.815)
County-, state-time dummies	*X*	*X*	*X*	*X*	*X*	*X*	*X*
Time-varying controls and interactions			*X*			*X*	*X*
Elev. instrument		*X*	*X*		*X*	*X*	*X*
Mean d.v. (1996)	*56%*	*56%*	*56%*	*41%*	*41%*	*41%*	*55%*
R^2	0.658	-	-	0.647	-	-	-
First-stage F-statistic	-	1314	28.68	-	1314	31.52	28.68
Number of counties	3,105	3,105	3,105	3,105	3,105	3,105	3,105

Notes: Standard errors are clustered by county. The analysis is always based on a panel of four presidential election cycles (1996-2008). The first raw gives the name of the dependent variable used Turnout is the proportion of the voting age population who casts a vote in presidential elections. Democratic/Republican vote share is the share of votes in presidential elections going towards the Democratic/Republican candidate. The Instrument is the logarithm of county-level mean elevation interacted with an indicator for the first year of broadband (i.e. 2000). Time-varying controls and itneractions are as in Table 1.4. County-level fixed effects absorb the level of those variables. ** indicates significance at the 5 percent level., *** indicates significance at the 1 percent level, and "-" indicates that the corresponding statistic cannot be meaningfully calculated.

Table 1.6

Effect of broadband on individual campaign donations (in dollars)

	Full sample		Presidential election years		Non-presidential years		
	1994, 1996, 2000-2008		1996, 2000, 2004, 2008		1994, 2002, 2008		1994, 2002
Number of ISPs	12,287***	34,378***	46,977***	-251.4	17,618***	7,333***	14,237***
	(597.0)	(4,122)	(5,509)	(4,887)	(2,314)	(2,283)	(2,911)
Zip code-, state-time dummies	*X*	*X*	*X*	*X*	*X*	*X*	*X*
County-level controls				*X*		*X*	*X*
zip code level controls				*X*		*X*	*X*
Elev. instrument		*X*	*X*	*X*	*X*	*X*	*X*
R^2	0.096	-	-	-	-	-	-
First stage F-statistic	-	614.9	629	180.6	568.3	172.5	46.57
Time dimension	7	7	4	4	3	3	3
Number of zip codes	27,405	27,405	27,405	27,405	27,405	27,405	27,405

Notes: Standard errors are clustered by zip code. With the exception of zip code level controls, all variables are as defined in Table 1.4 or Table 1.5. Zip code controls include log income, log tax returns, and log population. Zip code time interactions are time dummies interacted with 1994 log population, 1994 log total income, 1994 log total tax returns, 1994 contributions. Zip code fixed effects absorb the level of those variables. For each analysis, the election cycles that are included are mentioned in the top raw. The analysis of presidential election years uses 1996 rather than 1994 as a base year for time interactions controls. ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level, “-” indicates that the corresponding statistic cannot be meaningfully calculated.

Table 1.7

Effect of broadband on campaign donations across election cycles / political causes (in dollars)

	Non-pres. years	Pres. years	Pres. elections
PACs:	3,588**	-1,977	-
	(1,678)	(3,207)	-
mean dep. var.	*$9,508*	*$5,770*	-
House in-state:	1,423**	-1,326*	-
	(686.0)	(766.2)	-
mean dep. var.	*$3,362*)	*$4,350*	-
House out-of-state:	215.0	108.9	-
	(234.5)	(317.5)	-
mean dep. var.	*$728*	*$927*	-
President:	-	-	4,872**
	-	-	(2,032)
mean dep. var.	-	-	*$2,164*
Democratic candidate:	5,075***		5,385***
	(1,824)		(1,448)
mean dep. var.	*$5,347*		*$706*
Republican candidate:	2,398*		332.5
	(1,339)		(901.2)
mean dep. var.	*$6,198*		*$1,415*

Notes: Standard errors are clustered by zip code. For each cell, the title of the raw/column indicates the subset of donations that are included. For each category of donations, I include the regression estimate on the ISPs measure for the full specification of Table 1.6. Below the regression estimate, I report the amount contributed in the average zip code for the mentioned category in 1994 (non-presidential election cycles) or 1996 (presidentail election cycles). * indicates significance at the 10 percent level, ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level, "-" indicates that the corresponding statistic cannot be meaningfully calculated.

Table 1.8

Heterogeneity of broadband effect across donor/donations characteristics (in dollars)

Female:	2,612**	Male:	4,182**
	(1,269)		(2,077)
mean dep. var.	*$3,007*		*$9,255*
Retired:	-24.55	Professionals:	2,250***
	(382.8)		(640.1)
mean dep. var.	*$910*		*$985*
≤ $250:	264.7***	≤ $1000:	1,953***
	(87.11)		(453.5)
mean dep. var.	*$925*		*$8,619*
>$10,000:	580.5	>$1,000:	3,166
	(2,396)		(3,536)
mean dep. var.	*$3,734*		*$5,500*

Notes: Standard errors are clustered by zip code. Data come from non-presidential election cycles (1994, 2002, 2008). Each raw lists mutually exclusive groups of donations based on donor/donation characteristics. The categories are not meant to be mutually exhaustive, because of missing data / data fields with open-ended questions. For each category of donations, I include the regression estimate on the ISPs measure for the full specification of Table 1.6. Below the regression estimate, I report the amount of each type of money contributed in 1994.

Table 1.9

Distributing of zip codes based on frequency of contributions

	0	1	2	3	4	5	6	7
# zip codes	1,488	1,521	1,582	1,696	1,753	2,146	2,834	14,385
prop. of donations (%)	0.0	0.2	0.3	0.6	1.1	2.0	4.0	91.8
mean population	*589*	*807*	*1,029*	*1,376*	*1,699*	*2,303*	*3,618*	*16,478*

Notes: The summary statistics come from the sample of 27,405 zip codes. I label each column with the number of election cycles a zip code's residents donate a positve amoutn to federal political causes. As the panel of elections between 1994 and 2008 (excluding 1998) is used, the maximum number of appearances is 7. The first raw reports the number of zip codes in each category. The second raw gives the percentage of donations emanating from each category of zip codes. The third raw gives the mean population of zip codes in each category using 2000 Census data.

Table 1.10

Robustness analysis for campaign donations analysis (in dollars)

	1994, 1998, 2002, 2006		1994, 2002, 2006
	Zero 1998 broadband	Interp. 1998 broadband	No zeros
Number of ISPs	8,465***	7,843***	7,496***
	(1,899)	(2,193)	(2,268)
First-stage F-statistic	130.28	130.28	185
Number of zip codes	27,405	27,405	25,917

Notes: Standard errors are clustered by zip code. I include the full list of controls from Table 1.6. The first raw mentions the election cycles used in the analysis. The full sample of zip codes is used in the first two columns; the last column excludes the 1,488 zip codes whose residents never contribute between 1994 and 2008 (excluding 1998). In the first two columns, I use impute a value of zero to broadband for 1998, and half of the 2000 value (respectively). *** indicates significance at the 1 percent level.

Table 1.11

Analysis of pre-broadband trends

	Turnout		Dem. vote share		Dollars
	1980-1992	1984-1996	1980-1992	1984-1996	1990, 1994
Number of ISPs	-1.29**	0.0760	-0.0340	-1.36***	1,418
	(0.594)	(0.608)	(0.471)	(0.446)	(1,575)
First-stage F-statistic	15.18	17.16	16.94	18.02	48.71
Number of zip codes	-	-	-	-	27,405
Number of counties	3,105	3,105	3,105	3,105	-

Notes: Standard errors are clustered by county, except the last column where they are clustered by zip code. The first raw indicates the dependent variable used. The election years mentioned refer to the time frame used for the dependent variables. The first four columns repeat the analysis from the full specifications of Table 1.5. The last column repeats the analysis of the last column of Table 1.6. In each case, I include the corresponding time interactions for the new dependent variables. ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level, "-" indicates that the corresponding statistic cannot be meaningfully calculated.

Table 1.12

Effect of broadband on political knowledge, moral values and medium for campaign donations

	Social values			Medium for donations		Pol. know.
	Pro-choice	Pro-gay	Liberal index	Online	Offline	
Number of ISPs	0.850***	0.957***	0.887***	0.654**	-0.0289	0.538***
	(0.180)	(0.110)	(0.166)	(0.272)	(0.448)	(0.107)
mean of d.v. (%)	*59*	*38*	*50*	*5.3*	*17.5*	*62*
Obs.	11,349	20,201	8,546	2,178	2,178	5,103
R^2	0.167	0.239	0.268	0.090	0.116	0.278

Notes: Depending on data availability, the regressor is the number of ISPs in either the zip code or county of the respondent. Standard errors are accordingly clustered at either the zip code or county level. The title of each column indicates the dependent variable used. The political knowledge and liberal indeces are as defined in the text. The other dependent variables are indicator values, rescaled to a 0 / 100 scale, for whether the behavior mentioned is exhibtied by the respondent. In all regressions, I control for the following respondent demographics: religion, age, party affiliation, employment status, gender, urban/suburban/rural area, income, race, educational attainment, home ownership, marital status, life satisfaction, household size, household head. I also include a set of dummies indicating a missing value for each one of these controls. *** indicates significance at the 1 percent level. ** indicates significance at the 5 percent level.

Table 1.13

Effect of broadband on medium used to consume political news

	General Media				Internet		Type of TV News	
	TV	Newsp.	Radio	Mag.	Avg. effect	Aud. sorting	CTV	Non-CTV
Number of ISPs	-0.652***	0.0751	-0.0436	-0.0415	1.58***		-0.376**	-0.474***
	(0.113)	(0.122)	(0.184)	(0.0891)	(0.108)	1.42***	(0.187)	(0.172)
# ISPs * Dem.						(0.148)		
						1.65***		
#ISPs * Rep.						(0.163)		
						1.47***		
#ISPs * Indep.						(0.273)		
						1.86***		
#ISPs * No party.						(0.173)		
mean of d.v. (%)	*76*	*35*	*15*	*3*	*16*	*16*	*43*	*40*
Obs.	18,969	18,969	18,969	18,969	18,969	18,969	9,529	9,529
R^2	0.081	0.070	0.039	0.021	0.166	0.166	0.016	0.061

Notes: Depending on data availability, the regressor is the number of ISPs in either the zip code or county of the respondent (see Table 1.10). Standard errors are accordingly clustered at either the zip code or county level. I control for the same respondent characteristics as in Table 1.12. The title of each column indicates the dependent variable used, each defined as an indicator value, rescaled to a 0 / 100 scale, for whether the media mentioned is used. *** indicates significance at the 1 percent level. ** indicates significance at the 5 percent level.

Table 1.14
Effect of broadband on TV channel used to consume political news

	MSNBC	CNN	Local	NBC	ABC	CBS	FOX
Number of ISPs	-0.123	-0.109	0.004	-0.106	-0.0783	-0.360***	-0.227**
	(0.147)	(0.158)	(0.130)	(0.115)	(0.117)	(0.0902)	(0.100)
mean of d.v. (%)	*21*	*21*	*15*	*13*	*12*	*10*	*7*
Obs.	9,529	9,529	9,529	9,529	9,529	9,529	9,529
R^2	0.055	0.039	0.020	0.019	0.023	0.028	0.010

Notes: Standard errors are clustered in the same way as in Table 1.13. In all regressions, I include the same controls, and define dependent variables, as in Table 1.13. *** indicates significance at the 1 percent level. ** indicates significance at the 5 percent level.

References

Ansolabehere, S., J. M. de Figueiredo, and J. M. Snyder Jr. "Why is There so Little Money in U.S. Politics?," *Journal of Economic Perspectives,* 17 (2003), 105–130.

Aron, Debra, and David Burnstein. "Broadband adoption in the United States: An empirical analysis." Available at SSRN 386100 (2003).

Bartels, Larry M. "Uninformed votes: Information effects in presidential elections." *American Journal of Political Science* (1996): 194–230.

Bascom III, Earle C. Rusty, and Victor D. Antoniello. "Underground Power Cable Considerations: Alternatives to Overhead." (2011).

Battaglini, Marco, Rebecca B. Morton, and Thomas R. Palfrey. "The swing voter's curse in the laboratory." *Review of Economic Studies* 77, no. 1 (2010): 61–89.

Bednar, Jenna, and Elisabeth R. Gerber. "Political Geography, Campaign Contributions, and Representation." Unpublished manuscript (2012).

Bellou, Andriana. "The impact of Internet diffusion on marriage rates: Evidence from the broadband market." (2013).

Besley, Timothy, and Andrea Prat. "Handcuffs for the grabbing hand? The role of the media in political accountability." *American Economic Review* 96, no. 3 (2006): 720–736.

Bhuller, Manudeep, Tarjei Havnes, Edwin Leuven, and Magne Mogstad. "Broadband Internet: An Information Superhighway to Sex Crime." *Review of Economic Studies* (2013).

Bimber, Bruce. "Measuring the gender gap on the Internet." *Social Science Quarterly* 81, no. 3 (2000): 868–876.

Bimber, Bruce. "Information and political engagement in America: The search for effects of information technology at the individual level." *Political Research Quarterly* 54, no. 1 (2001): 53–67.

Bimber, Bruce Allen, and Richard Davis. *Campaigning online: The Internet in US elections.* New York: Oxford University Press, 2003.

Bresnahan, Timothy F., and Peter C. Reiss. "Entry and competition in concentrated markets." *Journal of Political Economy* (1991): 977–1009.

Brown, Jeffrey Robert, and Austan Goolsbee. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *Journal of Political Economy* 110, no. 3 (2002): 481–507.

Campante, Filipe R., Ruben Durante, and Francesco Sobbrio. "Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation". *NBER Working Paper* No. 19029 (2013).

Chiang, Chun-Fang, and Brian Knight. “Media bias and influence: Evidence from newspaper endorsements.” *Review of Economic Studies* 78, no. 3 (2011): 795–820.

Cox, Gary W., and Michael C. Munger. “Closeness, expenditures, and turnout in the 1982 US House elections.” *American Political Science Review* (1989): 217–231.

Crandall, Robert, Robert Hahn, Robert Litan, and Scott Wallsten. “Bandwidth for the People.” *Policy Review* 127 (2004): 04–01.

Czernich, Nina. “Broadband Internet and Political Participation: Evidence for Germany.” *Kyklos* 65, no. 1 (2012): 31–52.

Degan, Arianna, and Antonio Merlo. “Do voters vote ideologically?.” *Journal of Economic Theory* 144, no. 5 (2009): 1868–1894.

DellaVigna, Stefano, and Ethan Kaplan. “The Fox News effect: Media bias and voting.” *Quarterly Journal of Economics* 122, no. 3 (2007): 1187–1234.

DellaVigna, Stefano, and Matthew Gentzkow. “Persuasion: Empirical Evidence.” *Annual Review of Economics* 2, no. 1 (2010): 643–669.

DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel. “Persuasion bias, social influence, and unidimensional opinions.” *Quarterly Journal of Economics* 118, no. 3 (2003): 909–968.

Dettling, Lisa J. “Home Internet Use and Female Labor Supply.” Unpublished manuscript (2013).

Downs, Anthony. “An economic theory of democracy.” (1957): 260–276.

Durante, Ruben, and Brian Knight. “Partisan control, media bias, and viewer responses: Evidence from Berlusconi’s Italy.” *Journal of the European Economic Association 10*, no. 3 (2012): 451–481.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei. “Voting Alone? The Political and Cultural Consequences of Commercial TV.” Unpublished manuscript (2013).

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya. “Media and political persuasion: Evidence from Russia.” *American Economic Review* 101, no. 7 (2011): 3253–3285.

Erskine, Robert H., Timothy R. Green, Jorge A. Ramirez, and Lee H. MacDonald. “Digital elevation accuracy and grid cell size: Effects on estimated terrain attributes.” *Soil Science Society of America Journal* 71, no. 4 (2007): 1371–1380.

Eyster, Erik, and Matthew Rabin. “Rational and Naive Herding.” (2009).

Fabrikant, Geraldin, and Seth Schiesel. “Satellite vs. Cable: A Rivalry Beyond TV”, N.Y. TIMES, Feb. 19, 2001

Falck, Oliver, Robert Gold, and Stephan Heblich. "E-Lections: Voting Behavior and the Internet." Working Paper (2013).

Faulhaber, Gerald R. "Broadband Deployment: Is Policy in the Way?" *Broadband: Should We Regulate High-Speed Internet Access?* (2002): 223.

Feddersen, Timothy J., and Wolfgang Pesendorfer. "The swing voter's curse." *American Economic Review* (1996): 408–424.

Ferguson, Charles H. "The US Broadband Problem." *Brookings Review* 19, no. 1 (2001).

Flamm, Kenneth. "The role of economics, demographics, and state policy in broadband availability." *LBJ School of Public Affairs* (2005).

Freund, Caroline L., and Diana Weinhold. "The effect of the Internet on international trade." *Journal of International Economics* 62, no. 1 (2004): 171–189.

Gentzkow, Matthew. "Television and voter turnout." *Quarterly Journal of Economics* 121, no. 3 (2006): 931–972.

Gentzkow, Matthew, and Jesse M. Shapiro. "Media Bias and Reputation." *Journal of Political Economy* 114, no. 2 (2006): 280–316.

Gentzkow, Matthew, and Jesse M. Shapiro. "What drives media slant? Evidence from US daily newspapers." *Econometrica* 78, no. 1 (2010): 35–71.

Gentzkow, Matthew, and Jesse M. Shapiro. "Ideological segregation online and offline." *Quarterly Journal of Economics* 126, no. 4 (2011): 1799–1839.

Gerber, Alan, Dean Karlan, and Daniel Bergan. "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions." *American Economic Journal: Applied Economics* 1 (2009), no. 2: 35–52.

Gillett, Sharon E., William H. Lehr, and Carlos Osorio. "Local government broadband initiatives." *Telecommunications Policy* 28, no. 7 (2004): 537–558.

Gimpel, James G., Frances E. Lee, and Shanna Pearson-Merkowitz. "The check is in the mail: Interdistrict funding flows in congressional elections." *American Journal of Political Science* 52, no. 2 (2008): 373–394.

Golde, S. D. and N. H. Nie. "The effect of online news on political behavior." Unpublished manuscript (2010)

Goolsbee, Austan. "Subsidies, the value of broadband, and the importance of fixed costs." *Broadband: Should we regulate high-speed Internet access* (2002): 278–294.

Goolsbee, Austan, and Peter J. Klenow. "Evidence of learning and network externalities in the diffusion of home computers." *Journal of Law and Economics* 45, no. 2 (2002): 317–343.

Goolsbee, Austan, Michael F. Lovenheim, and Joel Slemrod. "Playing with fire: Cigarettes, taxes, and competition from the Internet." *American Economic Journal: Economic Policy* (2010): 131–154.

Greenstein, Shane M. "The economic geography of Internet infrastructure in the United States". Working Paper (2004).

Greenstein, Shane, Susan McMaster, and Pablo T. Spiller. "The effect of incentive regulation on infrastructure modernization: Local exchange companies' deployment of digital technology." *Journal of Economics & Management Strategy* 4, no. 2 (1995): 187–236.

Greenstein, Shane, and Jeff Prince. "The Diffusion of the Internet and the Geography of the Digital Divide in the United States." *NBER Working Paper* No. 12182 (2006).

Grenzke, Janet. "Comparing contributions to US House members from outside their districts." *Legislative Studies Quarterly* (1988): 83–103.

Grubesic, Tony H. "A spatial taxonomy of broadband regions in the United States." *Information Economics and Policy* 18, no. 4 (2006): 423–448.

Grubesic, Tony H., and Alan T. Murray. "Waiting for broadband: Local competition and the spatial distribution of advanced telecommunication services in the United States." *Growth and Change* 35, no. 2 (2004): 139–165.

Hazlett, Thomas W., and George Bittlingmayer. "The Political Economy of Cable Open Access". *Stanford Technology Law Review* (2003): 4–5.

Hazlett, Thomas W., and Arthur M. Havenner. "The arbitrage mirage: Regulated access prices with free entry in local telecommunications markets." *Review of Network Economics* 2, no. 4 (2003).

Hazlett, Thomas W., and Anil Caliskan. "Natural experiments in US broadband regulation." *Review of Network Economics* 7, no. 4 (2008).

Hindman, Douglas Blanks. "The rural-urban digital divide." *Journalism & Mass Communication Quarterly* 77, no. 3 (2000): 549–560.

Hindman, Matthew. "The real lessons of Howard Dean: Reflections on the first digital campaign." *Perspectives on Politics* 3, no. 1 (2005): 121–128.

Horton, Radley, Vivien Gornitz, Malcolm Bowman, and Reginald Blake. "Climate observations and projections." *Annals of the New York Academy of Sciences* 1196, no. 1 (2010): 41–62.

Jimeniz, Ed, and Shane Greenstein. "The emerging Internet retailing market as a nested diffusion process." *International Journal of Innovation Management* 2, no. 03 (1998): 281–308.

Kaid, Lynda Lee, Monica Postelnicu, Kristen Landreville, Hyun Jung Yun, and Abby Gail LeGrange. "The effects of political advertising on young voters." *American Behavioral Scientist* 50, no. 9 (2007): 1137–1151.

Kahn, Alfred E., Timothy J. Tardiff, and Dennis L. Weisman. "The Telecommunications Act at three years: an economic evaluation of its implementation by the Federal Communications Commission." *Information Economics and Policy* 11, no. 4 (1999): 319–365.

Landry, Craig E., and Mohammad R. Jahan-Parvar. "Flood Insurance Coverage in the Coastal Zone." *Journal of Risk and Insurance* 78, no. 2 (2011): 361–388.

Larcinese, Valentino. "Does political knowledge increase turnout? Evidence from the 1997 British general election." *Public Choice* 131, no. 3-4 (2007): 387–411.

Larcinese, Valentino, and Luke Miner. "The Political Impact of the Internet on US Presidential Elections." Unpublished manuscript (2012).

Lassen, David Dreyer. "The effect of information on voter turnout: Evidence from a natural experiment." *American Journal of Political Science* 49, no. 1 (2005): 103–118.

Lessig, Lawrence. *Code: And other laws of cyberspace.* Basic Books (AZ), 1999.

Lessig, Lawrence. *Republic, Lost: How Money Corrupts Congress–and a Plan to Stop It.* Hachette Digital, Inc., 2011.

Liang, Che-Yuan, and Mattias Nordin. "The Internet, News Consumption, and Political Attitudes–Evidence for Sweden." *BE Journal of Economic Analysis & Policy* 13, no. 2 (2013): 1071–1093.

Lizzeri, Alessandro, and Nicola Persico. "The provision of public goods under alternative electoral incentives." *American Economic Review* (2001): 225–239.

Matsusaka, John G. "Explaining voter turnout patterns: An information theory." *Public Choice* 84, no. 1-2 (1995): 91–117.

Michel-Kerjan, Erwann O., and Carolyn Kousky. "Come rain or shine: Evidence on flood insurance purchases in Florida." *Journal of Risk and Insurance* 77, no. 2 (2010): 369–397.

Miner, Luke. "The unintended consequences of Internet diffusion: Evidence from Malaysia". Working Paper (2013).

Moore, Ian Donald, R. B. Grayson, and A. R. Ladson. "Digital terrain modelling: a review of hydrological, geomorphological, and biological applications." *Hydrological processes* 5, no. 1 (1991): 3–30.

Moorthy, Satish Kumar. "The US cable television industry: the multi-service operator organizational structure as a bundle of competencies." PhD diss., Massachusetts Institute of Technology, 2009

Mullainathan, Sendhil, and Andrei Shleifer. "The market for news." *American Economic Review* (2005): 1031–1053.

Nordhaus, William D. "The economics of hurricanes and implications of global warming." *Climate Change Economics* 1, no. 01 (2010): 1–20.

Nie, Norman H., Darwin W. Miller III, Saar Golde, Daniel M. Butler, and Kenneth Winneg. "The world wide web and the US political news market." *American Journal of Political Science* 54, no. 2 (2010): 428–439.

Prat, Andrea, and David Strömberg. Commercial television and voter information. Centre for Economic Policy Research, 2005.

Prat, Andrea, and D. Strömberg. "The Political Economy of Mass Media," Working Paper (2011).

Prysby, Charles L., and B. lssue Voting. "Electoral behavior." *Annual Review of Political Science*, II (1987): 111–136.

Puglisi, Riccardo, and James M. Snyder Jr. "The Balanced U.S. Press." *NBER Working Paper* No. 17263, 2011.

Robinson, Glen O., and Dennis L. Weisman. "Designing competition policy for telecommunications." *Review of Network Economics* 7, no. 4 (2008).

Rosenzweig, Cynthia, William D. Solecki, Reginald Blake, Malcolm Bowman, Craig Faris, Vivien Gornitz, Radley Horton et al. "Developing coastal adaptation to climate change in the New York City infrastructure-shed: process, approach, tools, and strategies." *Climatic Change* 106, no. 1 (2011): 93–127.

Sinai, Todd, and Joel Waldfogel. "Geography and the Internet: Is the Internet a Substitute or a Complement for Cities?." *Journal of Urban Economics* 56, no. 1 (2004): 1–24.

Snyder Jr, James M., and David Strömberg. "Press Coverage and Political Accountability." *Journal of Political Economy* 118, no. 2 (2010): 355–408.

Starks, Michael. The Digital Television Revolution: Origins to Outcomes. Palgrave Macmillan, 2013

Strömberg, David. "Radio's impact on public spending." *Quarterly Journal of Economics* 119, no. 1 (2004): 189–221.

Strömberg, David. "How the Electoral College influences campaigns and policy: the probability of being Florida." *American Economic Review* 98, no. 3 (2008): 769–807.

Sunstein, Cass R. *Republic. com 2.0.* Princeton University Press, 2007.

Tedesco, John C. "Political information efficacy and Internet effects in the 2008 US presidential election." *American Behavioral Scientist* 55, no. 6 (2011): 696–713.

Tolbert, Caroline J., and Ramona S. McNeal. "Unraveling the effects of the Internet on political participation?." *Political Research Quarterly* 56, no. 2 (2003): 175–185.

Vaccari, Cristian. "From the air to the ground: The Internet in the 2004 US presidential campaign." *New media & society* 10, no. 4 (2008): 647–665.

Wallsten, Scott J. "Broadband penetration: an empirical analysis of state and federal policies." (2005).

Whitty, Monica T., and Deborah McLaughlin. "Online recreation: The relationship between loneliness, Internet self-efficacy and the use of the Internet for entertainment purposes." *Computers in Human Behavior* 23, no. 3 (2007): 1435–1446.

Willmott, Cort J., and Kenji Matsuura. "Smart interpolation of annually averaged air temperature in the United States." *Journal of Applied Meteorology* 34, no. 12 (1995): 2577–2586.

Xiao, Mo, and P. Orazem. "Is your neighbor your enemy? Strategic entry into the US broadband market". Working Paper (2009).

Xiao, Mo, and Peter F. Orazem. "Does the fourth entrant make any difference?: Entry and competition in the early US broadband market." *International Journal of Industrial Organization* 29, no. 5 (2011): 547–561.

Zimmerman, Rae, and Craig Faris. "Infrastructure impacts and adaptation challenges." *Annals of the New York Academy of Sciences* 1196, no. 1 (2010): 63–86.

Chapter 2
Hindsight-biased Courts and the Efficiency of the Legal System

2.1 Introduction

In the process of making sense of the complex world we live in, we routinely try to assign blame where it is due. We have even devised institutions – courts – that are entirely devoted to holding wrongdoers accountable for their actions. The legal system regulates a substantial share of economic activity, and a large literature has been devoted to assessing whether judges and juries are able to retrospectively process information. For courts to correctly assign blame, they need to first understand what information was available to decision-makers *at the time of decision-making*. In their investigations, legal scholars have documented a systematic tendency to overestimate the likelihood of an event after it happens, which psychologists have called *hindsight bias* (e.g. Fischhoff 1975, Rachlinski 1998). Hindsight bias interferes with courts' blame calculus by creating a false sense of inevitability. In this chapter, I fill an important gap in the literature by formally drawing out the implications of hindsight bias on the legal system.

The reliance on courts reduces the transaction costs involved with writing detailed contractual clauses about unlikely future events. If individuals correctly factor in the possibility of future liability at the time of decision-making, the legal system can be designed in a way to promote the social good. However, by design the legal system only deals with contentious cases, e.g. a tort suit only takes place in the event of an accident. A hindsight-biased court may then incorrectly utilize extraneous information it has at its disposal. Throughout, I assume that only the court is prone to the influence of hindsight bias, while other participants in legal proceedings anticipate the court's bias and act accordingly. I also draw an important distinction between situations where a court intervenes following a failure vs. success in risk management.

To study the implications of hindsight bias in a context where a judgment is cast following a failure in risk management, I consider the negligence standard. Courts apply the negligence standard to control accident damages, by requiring the compensation of victims when insufficient care was undertaken. Shavell (1987) defines accidents to be "harmful outcomes that neither injurers nor victims wished to occur – although either might have affected the *likelihood* or *severity* of the outcomes" [added emphasis]. This distinction between two types of preventive measures is clear in the context of automobile safety, where it is referred to as active safety and passive safety. One example of active safety is Brake Assist, a technology which helps increase braking pressure in an emergency situation.[55] An example of passive safety is seat belts, which reduce the impact of an accident.[56] I assume that the socially optimal investment involves installing high quality Brake Assist and low quality seat belts. In the classical analysis of negligence, the court correctly perceives welfare to only depend on *expected damage*, i.e. the mathematical product of the likelihood and severity of an accident. By requiring efficient effort on both dimensions, a rational court is then able to promote the social good.

It becomes indispensable to distinguish between active and passive safety in a world with a hindsight-biased court. A hindsight-biased court mistakenly overestimates the likelihood of an accident, which makes it attach too little value to Brake Assist but overly scrutinize seat belts. In anticipation, the car manufacturer may find it worthwhile to forego any investment in Brake Assist but install high quality seat belts. To shed additional light on the effect of hindsight bias, I then focus on the special case where a car manufacturer can only invest in passive safety. In this context, the socially efficient amount of investment in seat belts is increasing in the likelihood of the accident, as the investment becomes more likely to pay off. Hence the car manufacturer only (over)-invests into high quality seat belts when the likelihood of an accident exceeds a threshold. These findings closely conform with the recent trends in car manufacturing noted by Gladwell (2004):

> *Most of us think that S.U.V.s are much safer than sports cars. We feel that way because [...] in [a] TrailBlazer our chances of surviving a collision with a hypothetical tractor-trailer in the other lane are greater than they are in the Porsche. What we forget, though, is that in the TrailBlazer you're also much more likely to hit the tractor-trailer because you can't get out of the way in time. [...] The S.U.V. boom [...] [is] what happens when a larger number of drivers conclude [...] that they are better off treating accidents as inevitable rather than avoidable.*

To understand the implications of hindsight bias in a context where a judgment is cast following a success in risk management, I then turn my attention to the probable cause standard. A court applies the probable cause standard to uphold the Fourth Amendment's prescription about privacy,

[55] Other examples of active safety include good visibility from the vehicle, low interior noise levels, anti-lock braking system, and collision avoidance through automatic braking.

[56] Other examples of passive safety include air-bags, deformation zones, and laminated glass.

by invalidating evidence obtained through overly aggressive searches. A police force can increase the likelihood of uncovering incriminating evidence by engaging in more aggressive searches, and I assume it is in society's best interest to engage in highly aggressive searches. In the classical analysis of probable cause, the court correctly perceives welfare to only depend on *expected* crime, i.e. the mathematical product of the likelihood of a seizure and the disutility from crime. As in the case of negligence, a rational court can promote the social good by requiring the socially optimal of search effort. A hindsight-biased court then mistakenly overestimates the likelihood of a seizure, and hence does not recognize the responsiveness of crime control to high levels of search effort. In anticipation, the police does not invest in aggressive searches.

How can we remedy hindsight bias? One possibility is to encourage courts to think about what may have happened, i.e. engage in counterfactual thinking, a concept I now illustrate in the context of automobile safety. As will become clear, the court's counterfactual thoughts depend on both the manufacturer's investment in active safety (e.g. low quality Brake Assist), and the observed state of the world (e.g. an accident). I dub the "hard counterfactual" the possibility that no accident would have obtained with the same investment in active safety (i.e. low quality Brake Assist). The qualifier "hard" reflects the fact that the court needs to contradict a reality it has already observed. I then call "soft counterfactuals" possible outcomes (i.e. accident or no accident) for alternative levels of active safety (e.g. high quality Brake Assist). The qualifier "soft" reflects the fact that the court here contemplates worlds of a fairly hypothetical nature. In what follows, a "partially-biased" court overestimates the likelihood of an event for the *observed* level of active safety, but is able to think through soft counterfactuals.

In the negligence application, soft counterfactual thinking implies that biased misperceptions apply more broadly the greater the car manufacturer's investment in active safety. For example, when an accident materializes following a zero investment in Brake assist, a partially-biased court correctly perceives the marginal value of low quality Brake Assist. In contrast, when an accident materializes following an investment in low quality Brake Assist, a partially-biased court perceives the accident to have been inevitable for low quality Brake Assist. To evade liability, the car manufacturer then invests into low quality Brake Assist, a level of effort that lies between the (fully)-hindsight-biased and socially efficient levels. However, counterfactual thinking does not necessarily move the world toward efficiency, as the higher passive safety requirement may prove so costly as to discourage the car manufacturer to comply.

In the probable cause application, soft counterfactual thinking implies that biased misperceptions apply more broadly the *lower* the police force's search aggressiveness. For example, when a seizure follows no investment in aggressive searches, a partially-biased court perceives the seizure to have been inevitable for any level of search aggressiveness. In contrast, when a seizure follows a highly-aggressive search, a partially-biased court perceives the seizure to only have been inevitable for searches that are at least that aggressive. To be awarded a conviction, the police can now invest in any level of crime control weakly below the efficient level. Since the police only cares about crime control, it therefore chooses exactly the efficient level of search aggressiveness. This finding conforms with puzzling experimental evidence, according to which hindsight clouds judges' abilities to assess the likely outcome of the search, but does not influence their legal judgments (Wistrich et al. 2005, Rachlinski et al. 2011). In contrast with the negligence application, I conclude that counterfactual thinking is unambiguously socially beneficial in the probable cause context.

I contribute to the behavioral economics literature by characterizing the effect of a well-established cognitive deficiency. In particular, I add to a small formal literature on the effect of hindsight bias, which has so far focused exclusively on principal-agent relationships (political agency: Schuett and Wagner 2011; defensive medicine: Madarász 2012). My analysis makes clear that hindsight bias requires a nuanced approach to legal reform. This has substantial practical implications, as the lack of formal treatment in the legal literature has led to an incomplete understanding of this important psychological phenomenon. For example, in an influential paper Jolls et al. (1998) summarized the prevailing position of the legal profession by stating that "we can offer clear prescriptions because the hindsight bias points in only one direction: over-deterrence".[57] Finally, my study suggests a direction for identifying the extent of the bias outside of the lab, by suggesting distinct predictions depending on a typology that is irrelevant in the absence of hindsight bias (i.e. failure vs. success in risk management; accident-prevention vs. damage-mitigation).

The remainder of this chapter is organized as follows. Section 2.2 presents a general model of court decision-making. Section 2.3 defines hindsight bias. Sections 2.4 and 2.5 consider in turn the effect of hindsight bias on the negligence and probable cause standards. Section 2.6 concludes.

[57] This prediction of over-deterrence has led legal scholars to suggest a "no liability" rule as a potential remedy to hindsight bias (e.g. Rachlinski 1998). Rachlinski points out to the "business judgment rule" as a real-world application of this insight. According to this form of liability, corporate officers only held accountable when "grossly negligent".

2.2 A general model of court decision-making

2.2.1 Background

Regulation of harmful activities takes two main forms. Ex-ante regulation relies on government agencies announcing in advance what safety standards are to be followed, while ex-post regulation utilizes courts to deal with issues as they arise. While these two forms of regulation are sometimes used independently, they are in most instances used jointly (see Shavell (1984,1987), Kolstad et al. (1990) for formal models). Ex-ante regulation is preferred when dealing with predictable events, when harm has the potential of exceeding the injurer's assets, when legal suits may not be brought forward, or where there is substantial uncertainty about courts' legal standards (e.g. pollution control). In contrast, ex-post regulation is more appropriate when dealing with events that are hard to specify in advance, in which case it becomes less costly to let individuals interact freely with one another until a contention actually arises (e.g. neighborly disputes).

In this chapter, I study a general model of court decisions that is applicable to the study of two important types of legal determinations: negligence and probable cause. In negligence hearings, a court punishes a decision-maker (e.g. an individual, a firm) that imposed external harm after taking insufficient precautions. In probable cause hearings, a court discards evidence gathered through a police seizure that inappropriately infringed on privacy. The commonality between these two legal standards is that they both require a court to choose appropriate behavior on the basis of information available to the decision-maker (e.g. the police force) *at the time of decision-making.* They however differ in that the court intervenes following a failure (in negligence cases) vs. a success (in probably cause cases) in risk management. While other types of legal rulings exist, a discussion of their relative advantages and disadvantages is beyond the scope of this chapter. For the purposes of the analysis, it is sufficient to mention that the class of legal rulings I focus on has substantial practical relevance (see Shavell 2007 for a survey).

2.2.2 Formal model

The model has two players, a decision-maker (DM; "she") and a court (C; "it"). The state of the world S is binary, where the social benefit of a "success" is always higher than that of a "failure". DM and C engage in a simple two-stage interaction. In the first stage, DM can take precautions to either reduce the likelihood that an activity generates harm, or reduce the potential harm generated by such an activity. In the second stage, C casts a ruling on DM on the basis of a predetermined legal standard. Importantly, DM casts a judgment *after* observing the state of the world S, a crucial feature of the model that will allow hindsight bias to affect the outcome of the interaction.[58] The formal structure of the game is as follows:

Stage 1: DM chooses effort level (x, y) on two dimensions of effort. I refer to $y \geq 0$ as the investment in **damage-mitigation**, which consists of measures that alleviate the harm from an accident. As for investment $x \geq 0$ in **accident-prevention**, it consists of precautions that reduce the likelihood that an activity generates harm $D(y)$. Taking damage-mitigation as the numeraire, I call the unit cost of accident-prevention q_x. Following DM's effort choices, the state of the world $S = s \in \{0, 1\}$ is realized according to the rule

$$S = \begin{cases} 0 & \text{with probability } p(x) \\ 1 & \text{with probability } 1 - p(x) \end{cases},$$

where $p(.)$ is a differentiable, decreasing and convex function ($p' < 0$, $p'' > 0$). $S = 0$ refers to a "failure" (e.g. an accident, a crime) and $S = 1$ to a "success" (e.g. no accident, a police seizure). Hence a higher level of effort x decreases the likelihood of a failure, but becomes less effective at doing so for higher levels of effort. I then call $D(y)$ the damage associated with a failure $S = 0$ given effort level y, and assume that $D(.)$ is a differentiable, decreasing and convex function ($D' < 0$, $D'' > 0$). Hence an additional unit of effort y decreases the severity of an accident, but becomes less effective at doing so for higher levels of effort.

[58] I ignore considerations that would discourage victims bringing suit forward. One example that the literature has considered extensively is the cost of litigation.

Stage 2: The legal standard in place determines whether the court intervenes following state of the world $S = 0$ (negligence), or $S = 1$ (probable cause). I leave more detailed descriptions of each legal standard to later sections of the chapter.

2.3 Hindsight bias

2.3.1 Background

Ignoring available information is a challenging mental exercise. After an uncertain event takes place, individuals tend to overestimate the (ex-ante) predictability of its occurrence. In the psychology literature, this phenomenon has been referred to as *hindsight bias* (Fischhoff, 1975). This shortcoming in probabilistic assessment has garnered considerable empirical support across domains, particularly in the context of the legal system (e.g. tort law: Kamin and Rachlinski 1995, LaBine and LaBine 1996; criminal law: Wistrich et al. 2005, Rachlinski et al. 2011). Importantly, hindsight bias manifests itself not only in the general populace that makes up juries, but also in judges (see Viscusi 1999 and Guthrie et al. 2001).

The evidence that hindsight bias affects judges is testament to the power of this psychological force, as judges are heavily trained in assessing decisions under risk. In fact, by their own admission judges are aware of the possibility of falling prey to hindsight bias.[59] Furthermore, the adversarial nature of the legal system ensures judges gain exposure to the arguments formulated by contending sides. Finally, judges have ample opportunities to learn through experience as cases move through the courts system.[60]

[59] The claim that judges are aware of the existence of hindsight bias is exemplified by a statement made by Judge Frank Easterbrook in *Carroll v. Otis Elevator Co*. 1990: "no matter how conscientious jurors may be, there is a bias in the system. Ex post claims are overvalued and technical arguments are discounted in the process of litigation."

[60] For example, whenever a judge's decision is overturned by an appellate court, this gives a signal for an inaccurate initial decision.

2.3.2 A formalization of hindsight bias

A rational court who observes the state of the world S realizes such information was not available to DM at the time of choosing effort (x, y). However, a fully hindsight-biased court, C_s, mistakenly projects its knowledge of the realization of the state of the world, $S = s$, onto its perception of DM's ex-ante information set.[61]

Definition 1 (Full hindsight bias): *When the state of the world is revealed to be $S = s \in \{0, 1\}$, C_s perceives outcome $S = s$ to have been ex-ante fully predictable.*

The psychology literature suggests that to fight hindsight bias, it it useful to encourage individuals to think through counterfactual outcomes. However, this literature implicitly focuses on only one type of counterfactual, which I refer to as the **hard counterfactual**. In the model I study, the hard counterfactual requires thinking about the potential outcome, $S = 1 - s$, of DM's *observed* investment in accident-prevention, $x = a$. By definition, the more the court is able to think through the hard counterfactual, the less it overestimates the predictability of the observed state of the world, $S = s$. However, thinking through the hard counterfactual is mentally onerous since it requires imagining a world that does not conform with an observed reality.

Another type of counterfactuals, which I refer to as **soft counterfactuals**, consists of thinking of the effect of DM's *potential* actions, i.e. $x = a' \neq a$, on the state of the world S. Importantly, soft counterfactuals do not interfere with the perceived inevitability of the observed reality, i.e. "$S = s$ follows $x = a$". For that reason, encouraging courts to think through soft counterfactuals – rather than the hard counterfactual – seems like a more promising approach to fighting hindsight bias. As a consistency requirement, I assume that soft counterfactual thinking implies that following a failure (success), a failure (success) is also perceived to be inevitable for effort levels $x < a$ ($x > a$). This is a minimal restriction, as it only requires the realization that more precautions can only help avert a failure, and that fewer precautions cannot make a success any more likely. In what follows, I call **partial hindsight bias** the belief that the observed state of the world, $S = s$, was predictable for the *observed level* of accident-prevention, $x = a$, augmented with the ability to think through

[61] Note that in the following definition, a hindsight-biased court's perceptions do not depend in any way on the true probability of an accident, i.e. $p(a)$ where $x = a$ is the investment in accident-prevention.

soft counterfactuals.

Definition 2A (Partial hindsight bias when $S=0$): *When the state of the world is revealed to be $S=0$, $C_{a,0}$ perceives outcome $S=0$ to have been ex-ante fully predictable for an investment in accident-prevention* below *the observed level, $x \leq a$, but correctly perceives the marginal value of accident-prevention, $\widetilde{x} > a$.*

Definition 2B (Partial hindsight bias when $S=1$): *When the state of the world is revealed to be $S=1$, $C_{a,1}$ perceives outcome $S=1$ to have been ex-ante fully predictable for an investment in accident-prevention* above *the observed level $x \geq a$, but correctly perceives the infra-marginal value of accident-prevention, $\widetilde{x} < a$.*

I then call $p_0(.;a)$ ($p_1(.;a)$) the function that embodies $C_{a,0}$'s ($C_{a,1}$'s) perceptions of the relationship between effort and the likelihood of a failure (respectively). By directly applying the definition of partial hindsight bias, I obtain:

$$p_0(x;a) \equiv \begin{cases} 1 & \text{if } x \leq a \\ p(x) + (1 - p(a)) & \text{if } x > a \end{cases} \tag{3}$$

and

$$p_1(x;a) \equiv \begin{cases} p(x) - p(a) & \text{if } x \leq a \\ 0 & \text{if } x > a \end{cases}, \tag{4}$$

which I then illustrate in Figure 1.2 by depicting alongside $p_0(x;a)$ and $p_1(x;a)$ for a given function $p(x)$.

2.3 The negligence standard

One objective of accident law is to deter firms from engaging in socially harmful activities in an irresponsible way (Brown 1973; Diamond 1974). In this section, I focus on cases decided through the "negligence standard", according to which a decision-maker pays damages to the victim of an accident *only when* she took too few precautions to prevent the occurrence and reduce the severity of the accident.[62] The corresponding minimal levels of effort required by the court are referred to as *due care*. Because of its practical importance, negligence determination has been studied extensively by law and economics scholars (see Shavell 1987 for a book-length review).

2.3.1 The classical negligence standard

In this subsection, I outline the motivation and main finding from the classical study of the negligence standard in the law and economics literature. In order to motivate the existence of courts in this context, it is useful to describe – and compare – the unregulated and efficient outcomes from the game described in Section 2.2.

When unregulated, DM exerts no effort since she only directly incurs the cost of precautions, i.e. she solves $\min_{(x,y)\geq 0} q_x x + y$. I refer to this outcome as the **unregulated benchmark**. In contrast, efficient effort requires taking into account both the cost of effort and the expected harm to society from the activity,

$$\min_{(x,y)\geq 0} q_x x + y + p(x) D(y). \tag{5}$$

If the optimization problem (5) is quasiconcave, the **efficient benchmark** (x^*, y^*) solves the following First Order Conditions (FOCs):

$$\begin{aligned} q_x + p'(x) D(y) &= 0, \text{ and} \\ 1 + p(x) D'(y) &= 0. \end{aligned}$$

Because the forthcoming analysis will rely heavily on graphical intuition, I illustrate the shape of

[62] Judge Learned Hand wrote: "If the probability be called P; the injury, L; and the burden, B; liability depends upon whether B is less than L multiplied by P: i.e., whether B < PL." Judge Hand's statement corresponds to a discrete version of the model I describe in Section 2.2, as an investment whose cost is $B = q_x x + y$ either reduces the likelihood of an accident to $p(x) = 0$ or the damage associated with it to $D(y) = 0$.

the reaction curves in a quasiconcave problem (Figure 2.1). It is easy to see that the negative semi-definiteness of the Hessian at (x^*, y^*) – a necessary and sufficient condition for the quasiconcavity of problem (5) – implies that $\left|\frac{1}{x'(y^*)}\right| > \left|y'(x^*)\right|$.[63] In other words, $x(y)$ has to be steeper than $y(x)$ at the point at which they intersect (in the (x, y)-plane).

According to the negligence standard, following an accident $S = 0$ a (rational) court C requires DM to pay damages in the amount $D(y)$ to victims when the level of ex-ante precautions is below the socially efficient level. When indifferent between multiple value of effort, I assume that the court requires the smaller level of effort. Formally, DM solves the following problem in Stage 1 of the game:

$$\min_{(x,y)\geq 0} q_x x + y + p(x) D(y) \cdot I_{x<x^* \text{ or } y<y^*}, \tag{6}$$

where $I_.$ is the indicator function.

Since DM escapes liability by investing efficiently in precautions, she has no incentive to invest more than the efficient level. On the other hand, DM faces the social planner's problem (5) when effort is below the efficient level on either dimension. Conditional on taking effort below the efficient level, DM then finds it optimal to invest the socially optimal level of effort. By effectively creating a discontinuity in the firm's schedule of benefits at the efficient level of effort, the negligence standard therefore creates a sharp incentive for DM to act in society's best interest by choosing (x^*, y^*).

2.3.2 The negligence standard with a fully hindsight-biased court

Following an accident $S = 0$, a fully hindsight-biased court C_0 attaches a probability of 1 to an accident for all effort levels $x \geq 0$ on accident-prevention. As accident-prevention is perceived to be a costly investment with no benefit, C_0 sets due care on this dimension to 0. It then sets due care on damage-mitigation to a level

$$\bar{\bar{y}} \equiv \arg\min_{y\geq 0} y + D(y), \tag{7}$$

which it believes maximizes social welfare. In anticipation of C_0's negligence determination, DM's choice of effort depends on the relative cost of being held liable and escaping liability. Proposition

[63] The determinant of the second-order leading principal minor of the problem satisfies: $p(x) D(y) p''(x) D''(y) - (p'(x) D'(y))^2 > 0$. Furthermore, by total differentiation of the two FOCs we have $\frac{1}{x'(y)} = -\frac{p''(x(y))/p'(x(y))}{D'(y)/D(y)}$ and $y'(x) = \frac{-p'(x)/p(x)}{D''(y(x))/D'(y(x))}$. It is easy to see that the condition $\left|\frac{1}{x'(y^*)}\right| > \left|y'(x^*)\right|$ is equivalent to assuming that the necessary condition on the second-order leading principal minor is satisfied.

2.1 formalizes this discussion and makes precise the effect of full hindsight bias:

Proposition 2.1: *When the court is fully hindsight-biased, there are two potential outcomes from implementing the negligence standard:*

- *If* $q_x x^* + y^* + p(x^*) D(y^*) \leq \overline{\overline{y}}$, DM *chooses levels of effort* (x^*, y^*), C_0 *sets due care at* $\left(0, \overline{\overline{y}}\right)$, *and* DM *is held strictly liable for damages.*

- *If* $q_x x^* + y^* + p(x^*) D(y^*) > \overline{\overline{y}}$, DM *chooses levels of effort* $(0, \overline{y})$, C_0 *sets due care at* $\left(0, \overline{\overline{y}}\right)$, *and* DM *escapes liability.*

Proof of Proposition 2.1: In-text.

To escape liability, DM needs to invest $y = \overline{\overline{y}} > y^*$ in damage-mitigation. A fully hindsight-biased court C_0 requires a higher-than-efficient investment in damage-mitigation, as it perceives such ex-ante precautions to *always* pay off in the future. On the other hand, C_0 does not value accident-prevention, which discourages DM from exerting this type of effort. When the cost of evading liability is low enough, DM therefore invests too much in damage-mitigation, but too little in accident-prevention (from society's perspective).

This discussion leaves at least two important questions unanswered. What conditions need to be satisfied by the primitives of the model for full hindsight bias to lead to a distortion in DM's effort? How large is the potential welfare loss due to full hindsight bias? In the rest of this subsection, I shed light on theses issues by focusing on the case where DM can only invest in damage-mitigation (but not in accident-prevention).

Corollary 2.1: *Assume that* DM *cannot reduce the likelihood of an accident, i.e.* $p(x) = p$ *for all* x, *and denote* $y^*(p)$ *the corresponding socially optimal amount of damage-mitigation. Letting* $\overline{p} \equiv \left\{ p \middle| y^*(p) + pD(y^*(p)) = \overline{\overline{y}} \right\}$, *there are two potential outcomes from implementing the negligence standard when the court is fully hindsight-biased:*

- *If* $p \leq \overline{p}$, DM *chooses levels of effort* $y^*(p)$, C_0 *sets due care at* $\overline{y}$, *and* DM *is held strictly liable for damages.*

- *If* $p > \overline{p}$, DM *chooses levels of effort* $\overline{y}$, C_0 *sets due care at* $\overline{y}$, *and* DM *is never held liable.*

Proof of Corollary 2.1: In-text.

The full cost of efficient precautions, $y^*(p) + pD(y^*(p))$, increases with the likelihood p of an accident, as it becomes more likely victims have to be compensated.[64] Hence the more likely the accident is, the more attractive DM finds it to over-invest in damage-mitigation (i.e. choose $y = \overline{y}$) to escape liability. In Figure 2.3, I illustrate graphically that DM only finds it worthwhile to invest enough to evade liability for likely accidents, i.e. where $p > \overline{p}$.

Corollary 2.1 implies that full hindsight bias only creates a distortion when C_0's damage-mitigation requirement $\overline{y}$ is close enough to the efficient level $y^*(p)$. In fact, efficient effort increases with the likelihood of an accident, as it becomes more likely to reap off the benefit of damage-mitigation ($\frac{dy^*(p)}{dp} = \frac{-D'(\overline{y})}{pD''(\overline{y})} > 0$). In Corollary 2.2, I show that there is however no *a priori* indication that full hindsight bias has a bounded effect on welfare. Intuitively, this rests on the fact that welfare depends on both damage-mitigation, y, and the shape of the damage function, $D(.)$:

Corollary 2.2: *Assume that* DM *cannot reduce the likelihood of an accident, i.e.* $p(x) = p$ *for all* x, *and that the damage function takes the form* $D(y) \equiv \frac{\alpha}{y+1}$, *where* $\alpha \geq 0$. *When* $p \geq \overline{p} = \frac{1}{4}$, *the welfare loss from full hindsight bias becomes infinite as* $\alpha \to \infty$.

Proof of Corollary 2.2: See the Proofs Section.

[64] We have $\frac{d}{dp}\left[q_y y^*(p) + pD(y^*(p))\right] = D(y^*(p)) > 0$ by the envelope theorem.

2.3.3 The negligence standard with a partially hindsight-biased court

Following an accident $S = 0$, a partially biased court $C_{0;a}$'s perception of soft counterfactuals depends crucially on the level of effort DM exerted on accident-prevention, $x = a$. This makes DM's problem – in anticipation – fundamentally more complicated than under a fully biased court. According to the negligence standard, $C_{0;a}$ exonerates DM from liability if she exerted minimum levels of effort that simultaneously satisfy the following two equations:

$$\begin{aligned} x_0(y_0; a) &\equiv \min_{x' \geq 0} q_x x' + p_0\left(x'; a\right) D(y_0) \\ y_0(x_0; a) &\equiv \min_{y' \geq 0} y' + p_0(x_0; a) D\left(y'\right) \end{aligned}$$

Importantly, DM's choice of effort on the damage-mitigation dimension y is irrelevant for $C_{0;a}$'s due care calculus. In contrast, the observed level of accident-prevention is a crucial component of $C_{0;a}$'s perceptions because it directly affects the construction of soft counterfactuals through $p_0(.; a)$. Since the specification of partial hindsight bias assumes that $p_0(.; a)$ is differentiable everywhere *except at* $x = a$, $C_{0;a}$ must follow the following steps to determine due care on accident-prevention:

- Step 1: Determine the local optimum to the left of the observed level of effort, i.e. conditional on $x \leq a$.
- Step 2: Determine the local optimum to the right of the observed level of effort, i.e. conditional on $x > a$.
- Step 3: Determine the global optimum by comparing (perceived) utility at the two local optima from Steps 1 and 2 above.

In Proposition 2.2, I state the implications of partial hindsight bias on both the decision-maker and court's actions:

Proposition 2.2: *Let* $\underline{x} \equiv \left\{ a \middle| \overline{y}(a) = \overline{\overline{y}} \right\} \in \left(0, x\left(\overline{\overline{y}}\right)\right)$*, where* $x\left(\overline{\overline{y}}\right) < x^{*}$*. When the court is partially hindsight-biased, there are two potential outcomes from implementing the negligence standard:*

- *If* $q_x x^* + y^* + p(x^*) D(y^*) \leq q_x \underline{x} + \overline{\overline{y}}$, DM *chooses levels of effort* (x^*, y^*)*,* $C_{0;a}$ *sets due care at* $\left(0, \overline{\overline{y}}\right)$*, and* DM *is held strictly liable.*
- *If* $q_x x^* + y^* + p(x^*) D(y^*) > q_x \underline{x} + \overline{\overline{y}}$, DM *chooses levels of effort* $\left(\underline{x}, \overline{\overline{y}}\right)$*,* $C_{0;a}$ *sets due care at* $\left(0, \overline{\overline{y}}\right)$*, and* DM *is never held liable.*

Proof of Proposition 2.2: See the Proofs Section.

The basic logic for Proposition 2.2 is familiar from Proposition 2.1. However, it may be surprising that to evade liability, DM chooses a level of accident-prevention $\underline{x}$ that is *higher* than the level of due care that $C_{0;a}$ requires, $x = 0$. To build intuition, I discuss the special case where DM engages in an activity whose potential harm lies outside of its control, i.e. $D(y) = h > 0$ for all y. However, I still allow accident-prevention x to reduce the likelihood of an accident. If an accident $S = 0$ takes place after effort $x = a$, $C_{0;a}$ attaches a probability of $p_0(.; a)$ to the accident, and so sets due care at

$$\underline{x}(a) \equiv \min_{x \geq 0} q_x x + p_0(x; a) h.$$

Lemma 2.1 states the implications of partial hindsight bias in this simpler context:

Lemma 2.1: *Assume that the harm from an accident is fixed at a predetermined level* $D(y) = h > 0$ *for all* y*. Denote* $x^*(h)$ *the socially efficient level given a level of harm* h*, and let* $\underline{x}(h) \equiv \{ a | h = q_x x^*(h) + p_0(x^*(h); a) h\} \in (0, x^*(h))$*. Then the optimal level of accident-prevention perceived by a partially hindsight-biased court* $C_{0;a}$ *satisfies*

$$x_0(a) = \begin{cases} x^*(h) & \text{if } a < \underline{x}(h) \\ 0 & \text{if } a \geq \underline{x}(h) \end{cases},$$

where $x = a$ *is the level of accident-prevention exerted by* DM*. We then obtain the follow outcome from implementing the negligence standard:* DM *chooses effort* $\underline{x}(h)$*,* $C_{0;a}$ *sets due care at* 0*, and* DM *escapes liability.*

Proof of Lemma 2.1: See the Proofs Section.

When DM's investment in accident-prevention, $x = a$, is below a threshold, $a < \underline{x}(h)$, correct perceptions about the marginal values of effort apply more broadly than biased perceptions (and so $C_{0;a}$ sets due care at the efficient level). Conversely, when observed effort is above that threshold, biased perceptions about the inevitability of an accident apply to a larger range of effort levels (and so $C_{0;a}$ finds it futile for DM to invest in accident-prevention).

Given the definition of $p_0(.;a)$, one can then easily argue by contradiction that $\underline{x}(h) \in (0, x^*(h))$ (see Figure 2.3). When $a = 0$, $C_{0;a}$ correctly perceives the marginal value of additional accident-prevention, and so sets due care at $x^*(h)$. In contrast, when $a = x^*(h)$, $C_{0;a}$ misperceives the accident to be inevitable for $x \leq x^*(h)$, and so sets due care at 0. This establishes that the value of effort at which $C_{0;a}$ is indifferent between null and efficient accident-prevention lies between these two extremes.

What obtains when the negligence standard is implemented? Clearly, DM chooses to evade liability by investing the minimal possible amount in accident-prevention, $a = \underline{x}$. Doing so is more advantageous than being held liable for possible damages, in which case DM would be facing the social planner's problem. In fact, DM would in that case incur both the cost of efficient effort, $x^*(h) > \underline{x}(h)$, and the associated expected damages, $p(x^*(h))\,h$. I end this discussion by noting one important distinction between the two situations studied in Lemma 2.1 and Proposition 2.2: DM always underinvests in accident-prevention in the former case, but only sometimes does in the latter. This is because the court sets a stringent due care requirement on damage-mitigation (in Proposition 2.2), which may make DM unwilling to comply with the court's requirements.

2.3.4 Discussion

Under a system of strict liability, DM is *always* responsible for the compensation of victims for damages $D(y)$. It is clear that hindsight bias increases the relative attractiveness of the strict liability rule. That being said, an implicit assumption of my modeling approach is that the negligence standard is preferred to the strict liability standard. While I do not explicitly model why the negligence rule dominates strict liability, an element of negligence would be desirable whenever – as is almost always the case – multiple parties have a role to play in causing accidents, or when DM's assets are not sufficient to cover losses.[65] We should then attach a negative interpretation to the effect of hindsight bias when it turns the negligence rule into a strict liability rule (see Proposition 2.1, Corollary 2.1, and Proposition 2.2). As Shavell (1987) points out, it is unnatural to set due care at a level that no injurer would adhere to under the negligence rule, as this amounts to using strict liability. The question then becomes: do the relative benefits of negligence still outweigh the loss in efficiency caused by hindsight bias?

Would full hindsight bias have different implications if C_0 instead assigned a probability $\delta < 1$ to an accident? Assume that C_0 instead assigns a probability $\delta \in (p(0), 1)$ to accidents, where the lower bound reflects the idea that hindsight bias amounts to the overestimation of the likelihood of an accident. Then the statement of Proposition 2.1 would only require one change: replacing $\overline{y}$ by $\overline{y}_\delta \equiv \arg\min_{y \geq 0} y + \delta D(y)$. We have $\overline{y}_\delta < \overline{y}$, since the pay off to damage-mitigation increases with the likelihood of an accident. When $\delta \in (p(0), 1)$, it is then cheaper for – and therefore more likely that – DM complies with the hindsight-biased court's requirement, as due care is closer to the efficient level. Furthermore, social welfare is higher at $\overline{y}_\delta$ than $\overline{y}$ as this level of damage-mitigation more closely approximates $y(0)$, the optimal investment under zero accident-prevention.

How do the implications of partial hindsight bias differ from those of full hindsight bias? Both a partially and fully hindsight-biased court set due care at $(0, \overline{y})$; the difference in DM's behavior is entirely driven by her anticipation of $C_{0;a}$'s degree of soft counterfactual thinking. Since $C_{0;a}$'s counterfactual thinking requires DM to invest $\underline{x} > 0$ in accident-prevention to escape liability, partial hindsight bias causes a smaller distortion in effort than full hindsight bias. However, DM needs to invest on accident-prevention to evade liability under partial hindsight bias, which makes the negligence turn into strict liability more frequently than under full hindsight bias. On net, it is therefore unclear if soft counterfactual thinking is welfare-enhancing under the negligence standard.

[65] In the basic model that I present in this chapter, the negligence rule and strict liability rule have the same implications for DM's behavior. The equivalence between the two prevalent forms of liability however stops holding once the model is complicated (Shavell 1986).

2.4 The probable cause standard

The Fourth Amendment requires the police to weigh out the social goal of crime control and the desire to preserve privacy (see Packer 1964 for an early analysis of the criminal process).[66] In this section, I focus on one approach to meeting the Fourth Amendment's objectives, the "probable cause" standard. According to the probable cause standard, a court does not award the police the conviction of an individual – even if guilty of the crime he is accused of – when an arrest was the result of an overly-aggressive search.[67] The corresponding maximal level of search intensity allowed by the court is referred to as *exclusionary care.* The analysis in this Section follows the same general format as that of Section 4.

2.4.1 The classical probable cause standard

In this subsection, I outline the motivation and main findings from the classical study of the probable standard. To minimize the introduction of new notation, I call h the disutility to society from crime. The greater the level of search aggressiveness, x, the more likely a police is to make a seizure, $1 - p(x)$. Alternatively, $p(x)$ describes the payoff from aggressive searches through a reduction in the likelihood of crime. I think of $q_x x$ as the direct cost to society of aggressive searches in the form of privacy infringement. The police force (DM) bears no costs to aggressive searches but internalizes the benefit of crime control. To break ties, I however assume that when indifferent, DM chooses the lower level of effort.

It is easy to see that when unregulated, DM exerts infinite effort on minimizing crime, as her problem is $\min_{x \geq 0} p(x) h$. I refer to this outcome as the **unregulated benchmark**. In contrast, efficient effort requires taking into account both the cost of effort and the expected harm to society from the activity,

$$\min_{x \geq 0} q_x x + p(x) h. \tag{8}$$

[66]This idea has been expressed explicitly in legal rulings: "in judging reasonableness, we look to the gravity of the public concerns served by the seizure, the degree to which the seizure advances the public interest, and the severity of the interference with individual liberty" (*Illinois v. Lidster* 2004). Packer (1964) provides an early legal analysis of the criminal process. Note that this principle can also be enforced ex ante by requiring the police to obtain a warrant before performing a search.

[67]Legal rulings have expressed the idea in explicit terms: "in judging reasonableness, we look to the gravity of the public concerns served by the seizure, the degree to which the seizure advances the public interest, and the severity of the interference with individual liberty" (*Illinois v. Lidster* 2004). This principle can be enforced ex-ante by requiring the police to obtain a warrant before performing a search. The guiding principle is that warrants should be obtained in advance, unless expedite action is required, in which case the probable cause standard is applied following an arrest to decide whether evidence obtained can be included for prosecution.

The optimization problem (8) is quasiconcave since $p(.)$ is a decreasing and convex function, and so the **efficient benchmark** x^* solves the First Order Condition, $q_x + p'(x)h = 0$.

Following a police seizure $S = 1$, a (rational) court C decides whether to award a police force DM the conviction of a criminal. The legality of the search depends exclusively upon the state of knowledge of the police officers who performed it; the outcome of the search is thus irrelevant.[68] According to the probable cause standard, the conviction is only awarded when the level of search aggressiveness x was *below* the socially efficient level. Formally, DM solves the following problem in Stage 1 of the game:

$$\min_{x \geq 0} q_x x + p(x) h \cdot I_{x > x^*},$$

where $I_{.}$ is the indicator function.

Since DM escapes liability by investing efficiently in precautions, she has no incentive to invest more than the efficient level. On the other hand, DM faces the social planner's problem (5) when effort is *above* the efficient level on either dimension. Conditional on taking effort below the efficient level, DM then finds it optimal to invest the socially optimal level of effort. The crucial difference between the probable cause and negligence standards is that DM is now punished when *too much* effort was undertaken. In that case, the fruits of an illegal search are discarded, and so a criminal may escape prosecution even when guilty of the crime he is accused of.

2.4.2 The probable cause standard with a fully hindsight-biased court

Following a police seizure $S = 1$, a fully hindsight-biased court C_1 sets exclusionary care at $\underline{\underline{x}} \equiv \min_{x \geq 0} q_x x = 0$, which it perceives to maximize social welfare. Proposition 2.3 makes precise the effect of full hindsight bias:

Proposition 2.3: *When the court is fully hindsight-biased, the implementation of the probable cause standard leads to the following outcome:* DM *does not invest in crime-prevention,* C_1 *sets exclusionary care at* $\underline{\underline{x}} = 0$*, and* DM *is never awarded a conviction.*

Proof: In-text.

[68] See the legal rulings in *Weeks v. U.S.* 1916, *Map v. Ohio* 1961, *Wolf v. Colorado* 1979, *United States v. Leon* 1985.

The court believes that crime control would have been achieved regardless of whether an aggressive search was undertaken. But since the fully biased court C_1 correctly internalizes the cost of privacy breach, the police is only awarded a conviction when no crime-prevention is undertaken. Hence any level of search aggressiveness is perceived as being excessive. In anticipation of the court's behavior, DM does not engage in aggressive searches, which leads to too little crime control from a social perspective.

2.4.3 The probable cause standard with a partially hindsight-biased court

Assume that $C_{1;a}$ is a partially biased court that witnessed a police seizure $S = 1$ following an investment of $x = a$ in accident-prevention by DM. $C_{1;a}$'s perceptions about the likelihood of an accident are the mirror image of $C_{0;a}$'s perceptions (see Figure 2.2). This crucial distinction comes from the fact that a probable cause ruling follows a success $S = 1$ in risk management, while a negligence ruling follows a failure $S = 0$. Formally, $C_{1;a}$ awards DM a conviction if she exerts a maximal level of effort that solves

$$x_1(a) \equiv \min_{x \geq 0} q_x x + p_1(x; a) h.$$

In Proposition 2.4, I state the implications of partial hindsight bias on the probable cause standard:

Proposition 2.4: *Assume that the harm from crime is fixed at a predetermined level $D(y) = h > 0$ for all y. Then the optimal level of effort perceived by a partially hindsight-biased court $C_{1;a}$ satisfies*

$$x_1(a) = \min\{a, x^*\},$$

where $x = a$ is the level of crime-prevention exerted by DM. *We then obtain the following outcome from implementing the probable cause standard:* DM *invests the socially efficient level x^* in crime control, $C_{1;a}$ sets exclusionary care at x^*, and* DM *is awarded a conviction following a seizure.*

Proof: In-text.

The characterization of $p_1(.;a)$ implies that $C_{1;a}$ understands the infra-marginal value of crime-prevention, but never perceives the marginal value of additional effort. Hence when DM exerts at least the efficient level of effort, $C_{1;a}$ correctly perceives that the socially efficient level of crime control to maximize welfare. But when DM exerts less than the efficient level of effort, $C_{1;a}$ misperceives the observed level of crime-prevention to be optimal. DM is then awarded a conviction whenever effort is below the exclusionary threshold, i.e. for all a such that $a \leq x_1(a) = \min\{a, x^*\}$. This condition is clearly satisfied for any level of effort below the efficient level, $a \leq x^*$. Because DM always prefers having more crime control, she then chooses the socially efficient level of crime control. This finding conforms with recent experimental evidence, which suggests that judges 1) make similar rulings on probable cause in foresight and in hindsight, *but* 2) exhibit hindsight bias in their perception of the likely outcome of a search (Wistrich et al. 2005; Rachlinski et al. 2011). While puzzling at face value, such a duality follows naturally from the formalization of (partial) hindsight bias.

2.4.4 Discussion

Does the stark implication of full hindsight bias follow from the assumption that C_1 assigns a probability 1 to a seizure (Proposition 2.3)? Say that the court instead attaches a fixed probability γ to a seizure, then exclusionary care would instead have been set at $\underline{\underline{x}}_\gamma \equiv \min_{x \geq 0} q_x x + (1-\gamma) h = 0 = \underline{\underline{x}}$. Hence relaxing the definition of hindsight bias does not affect the conclusions in the text, unlike in the study of the negligence standard. Note that in the present context, the police force has no control over the level of crime. Presumably, the harm from crime, h, is determined by a criminal whose behavior lies outside of the model.

One upside is that under the probable cause standard, counterfactual thinking may arise more naturally, as the court may realize that it is not reasonable to expect that aggressive searches have absolutely no role to play in the outcome the police has come to court for. When this is the case, Proposition 2.4 says that partially biased courts are able to move society back to the socially efficient level of crime control, which implies that counterfactual thinking is an unambiguously effective remedy to hindsight bias in this context.

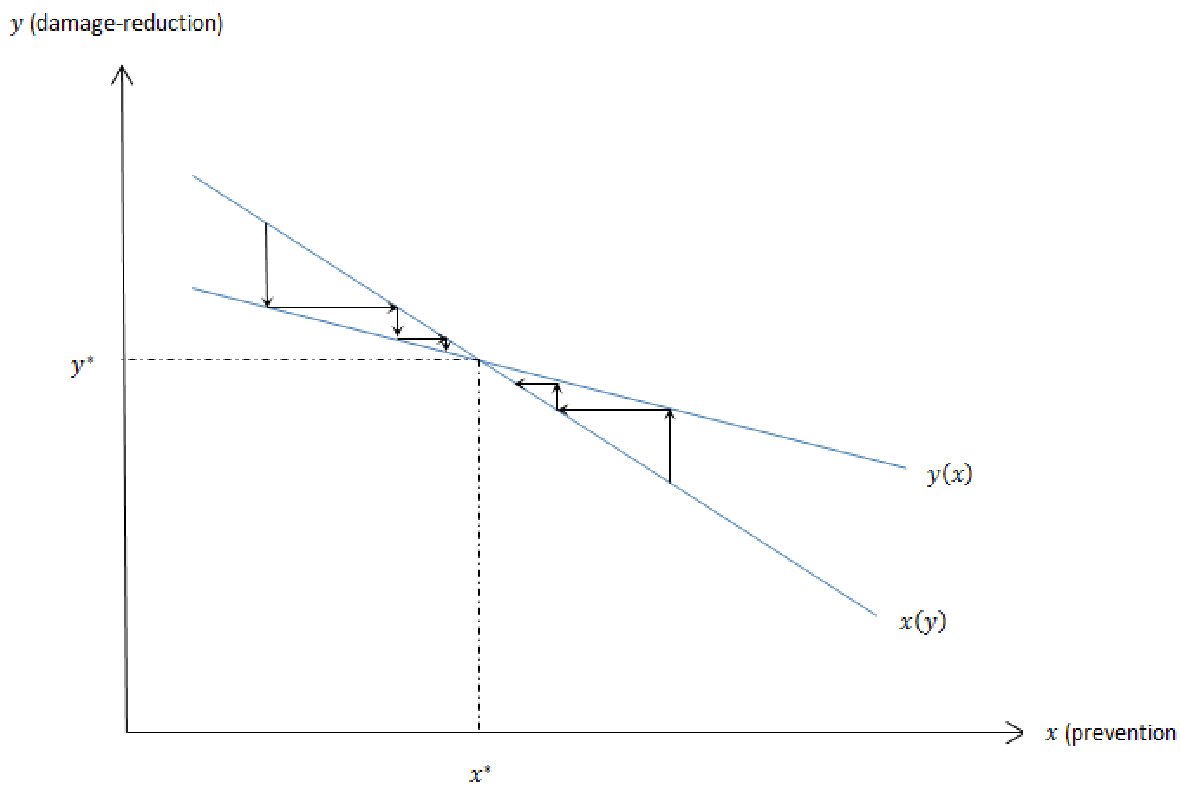
2.5 Conclusion

In this chapter, I have studied the implications of "hindsight bias" on two legal standards: negligence and probable cause. In contexts where the negligence rule is applied, I found that hindsight bias discourages preventive measures that reduce the likelihood of an accident, but encourages investment that reduces its severity. In contexts where the probable cause standard is applied, I showed that hindsight bias discourages aggressive searches that are meant to increase the likelihood of police seizures. I also studied the effectiveness of counterfactual thinking as a remedy for hindsight bias, and found that it is less effective under the negligence standard than the probable cause standard.

The present study of the effect of hindsight bias on the legal system can be extended in a number of directions. A natural next step would be to endogenize the occurrence of counterfactual thinking. Relatedly, the model does not allow a hindsight-biased court to learn from its past experiences. This is a particularly important limitation in the context of the negligence standard, where my findings suggest that a hindsight-biased court's perceptions do not conform with the decision-maker's actual choices. From a broader perspective, I have only investigated the implications of hindsight bias on the legal system. Moving forward, the hope is that the study of this psychological tendency will grow more commensurate with the enormous amount of evidence that has been produced for its existence.

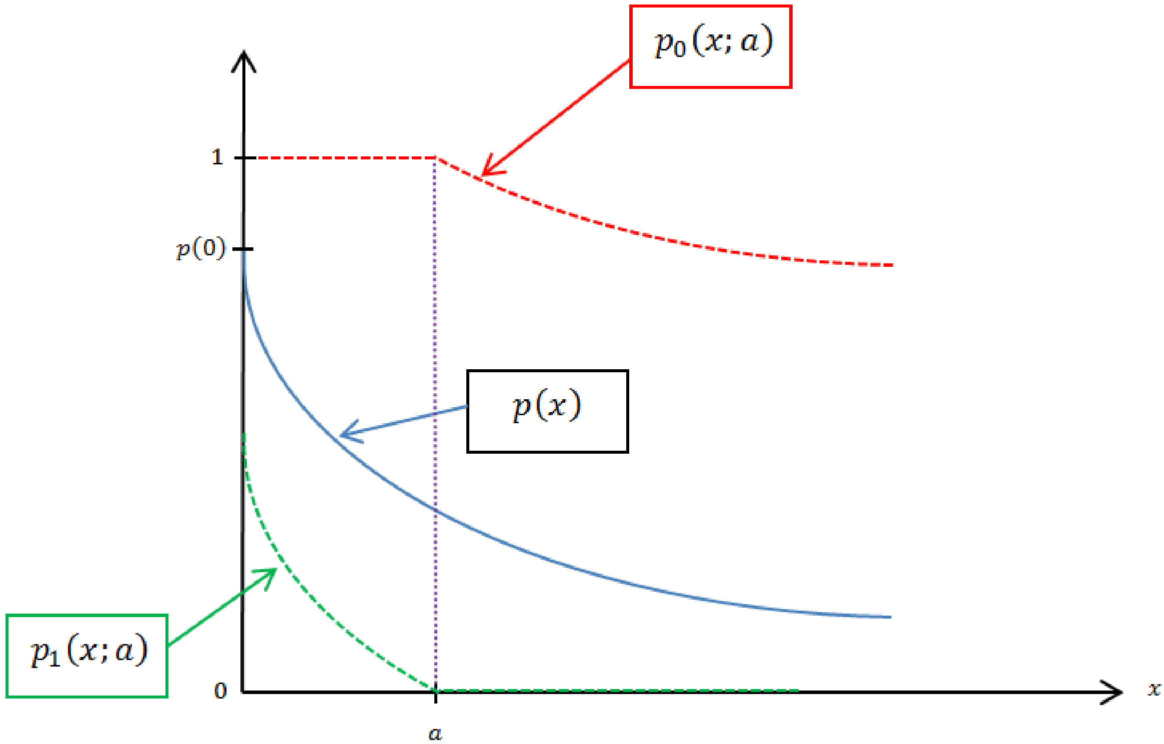
2.6 Figures

Figure 2.1: Optimal reaction curves



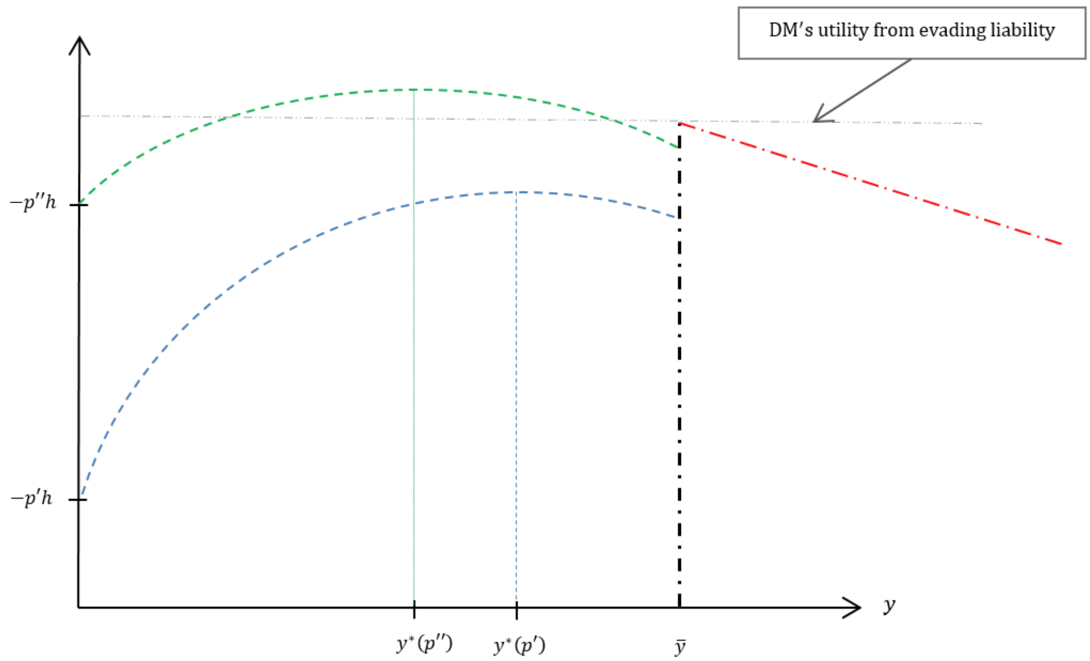
Notes: The decision-maker's optimal reaction curves for accident-prevention $(x(y))$ and damage-mitigation $(y(x))$. Starting with any level of effort $(x(y), y)$ on the $x(y)$ curve, a vertical arrow points toward a superior combination of effort levels $(x(y), y(x(y)))$ for DM. Starting with any level of effort $(x, y(x))$ on the $y(x)$ curve, a horizontal arrow points toward a superior combination of effort levels $(x(y(x)), y(x))$ for DM. This figure illustrates that whenever $x(y)$ is steeper than $y(x)$ in the $(x, y) - plane$, the intersection of the two curves corresponds with the socially efficient-level of effort, (x^*, y^*).

Figure 2.2: Partial hindsidght bias



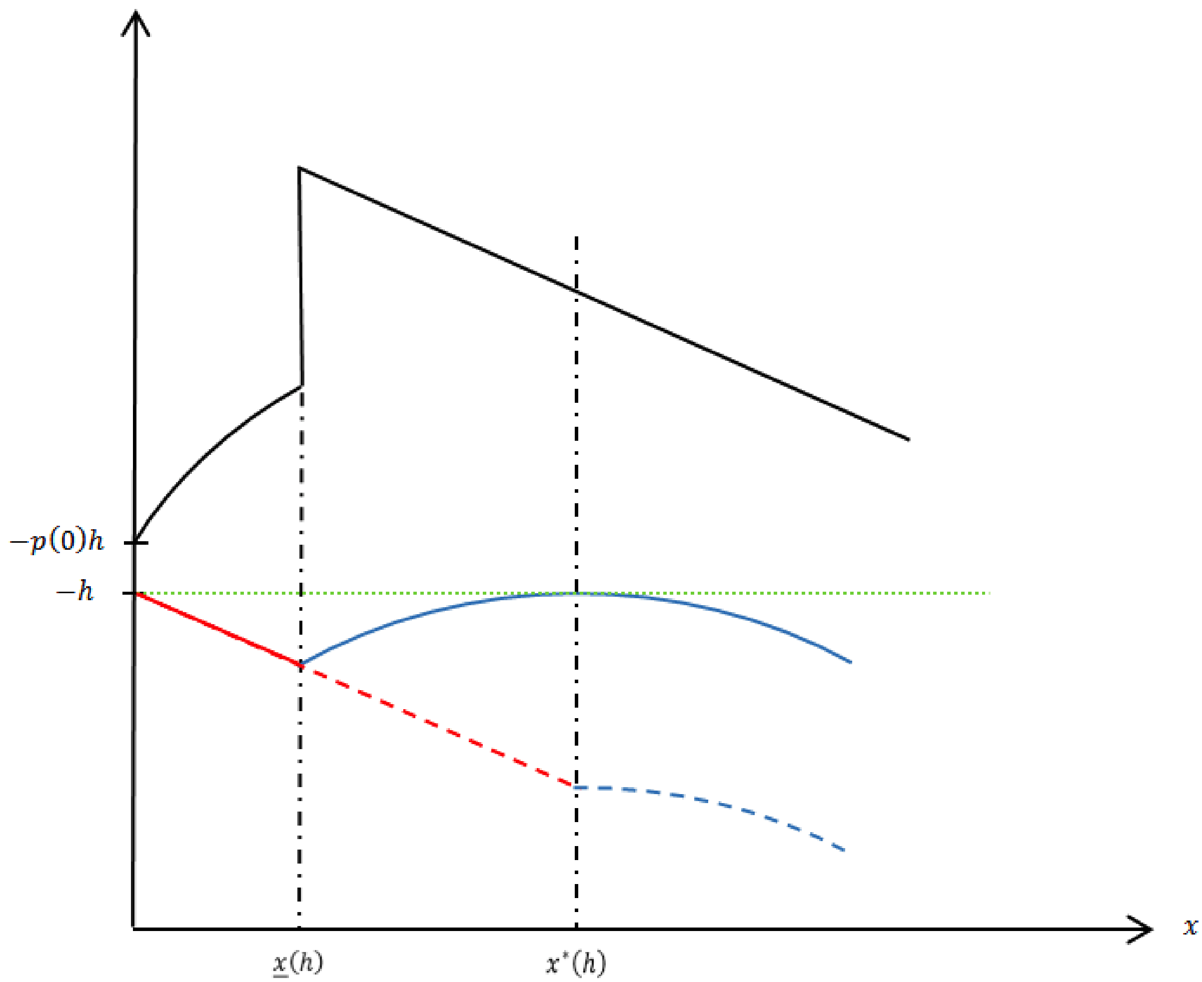
Notes: Illustration of partial hindsight bias given that DM exerts $x = a$ in accident-prevention. The blue solid line represents the likelihood of an accident for different levels of effort, $p(.)$. This curve highlights that 1) DM needs not fully control the outcome of the activity she engages in, 2) a failure does not necessarily follow when no effort is exerted ($p(0) < 1$), and 3) a success does not necessarily take place when maximal effort is undertaken ($\lim_{x\to\infty} p(x) > 0$). The red curve is $C_{0;a}$'s perception of the likelihood of an accident following $S = 0$ ($p_0(.;a)$), while the green line is $C_{1;a}$'s perception of the likelihood of an accident following a success $S = 1$ ($p_1(.;a)$.

Figure 2.3: The negligence standard under full hindsight bias



Notes: The negligence standard under full hindsight bias and an accident with a fixed probability p. This figure depicts DM's utility as a function of damage-mitigation y in two contexts: frequent accidents ($p = p' > \overline{p}$) and rare accidents ($p = p'' < \overline{p}$). The red line represents DM's utility from investing a minimum of $\overline{y}$ on damage-mitigation. Over this range, DM is not held liable for damages, and so only incur the direct cost of damage mitigation, y. The green (blue) curve represents DM's utility from internalizing the full social cost of the accident, which she must incur when exerting less than due care $y = \overline{y}$, when $p = p' > \overline{p}$ ($p = p'' < \overline{p}$).

Figure 2.4: The negligence under partial hindsight bias



Notes: The negligence standard under partial hindsight bias and a potential accident that causes a fixed damage h. The black curve depicts DM's utility as as a function of accident-prevention x. The Figure also illustrates the court's perception of DM's utility given two possible values of DM's investment in accident-prevention: $a = \underline{x}(h)$, and $a = x^*(h)$. I use blue curves to indicate DM's perceived utility for levels of accident-prevention above the observed level, $x = a$, where the court correctly perceives the marginal value of effort. I then use red curves to indicate DM's perceived utility for levels of effort below the observed level, $x = a$, where the court misperceives the accident to be inevitable regardless of the investment made in accident-prevention.

2.7 Proofs

Corollary 2.2: *Assume that* DM *cannot reduce the likelihood of an accident, i.e.* $p(x) = p$ *for all* x*, and that the damage function takes the form* $D(y) \equiv \frac{\alpha}{y+1}$*, where* $\alpha \geq 0$*. When* $p \geq \overline{p} = \frac{1}{4}$*, the welfare loss from full hindsight bias becomes infinite as* $\alpha \to \infty$*.*

Proof of Corollary 2.2: Solving for DM's optimal level of damage-mitigation, we obtain

$$\alpha p \left(\frac{1}{y^*(p) + 1} \right)^2 = 1,$$

and so $y^*(p) = \sqrt{\alpha p} - 1$, and $\overline{\overline{y}} = \sqrt{\alpha} - 1$. Plugging into the defining equation for $\overline{p}$, we obtain $\overline{p} \equiv \left\{ p | \sqrt{\alpha p} - 1 + p\frac{\alpha}{\sqrt{\alpha p}} = \sqrt{\alpha} - 1 \right\} = \left\{ p | 2\sqrt{\alpha p} = \sqrt{\alpha} \right\} = \frac{1}{4}$. The effect of full hindsight bias on welfare when $p \geq \overline{p}$ is then:

$$\begin{aligned} \overline{\overline{y}} - y^*(p) + p\left(D\left(\overline{y}\right) - D\left(y^*(p)\right)\right) &= \sqrt{\alpha}\left(1 - \sqrt{p}\right) + p\left(\frac{1}{\sqrt{\alpha}} - \frac{1}{\sqrt{\alpha p}} \right) \\ &\to \infty \text{ as } \alpha \to \infty \boxtimes \end{aligned}$$

Technical Lemma: *The reaction curves of* DM *perceived by a partially hindsight-biased court* $C_{0;a}$ *satisfy*

$$x_0(y_0; a) = \begin{cases} x(y_0) & \text{if } y_0 < \overline{y}(a) \\ 0 & \text{if } y_0 \geq \overline{y}(a) \end{cases}.$$

and

$$y_0(x_0; a) = \begin{cases} \overline{\overline{y}} & \text{if } x_0 \leq a \\ D'^{-1}\left(-\frac{1}{p(x_0)+1-p(a)} \right) & \text{if } x_0 > a \end{cases},$$

where $x = a$ *is the level of accident-prevention exerted by* DM*,* $x(.)$ *and* $y(.)$ *refer to* DM*'s correct reaction curves,* $\overline{y}(a) \equiv \{y | D(y) = q_x x(y) + p_0(x(y); a) D(y)\} < \{y | x(y) = a\}$ *and* $\overline{\overline{y}}$ *is defined as in (7).*

Proof of Technical Lemma: Starting with damage-mitigation, there are two cases to consider. When $x_0 \leq a$, $C_{0;a}$ perceives the accident to be inevitable ($p(x_0; a) = 1$) and so $y_0(x_0; a) = \overline{\overline{y}}$. When $x_0 > a$, $y_0(x_0; a)$ is implicitly defined by the FOC

$$q_y + p_0(x_0; a) D'(y_0(x_0; a)) = q_y + [p(x_0) + 1 - p(a)] D'(y_0(x_0; a)) = 0.$$

Comparing $y_0(x_0; a)$ to $y(x_0)$, we then have $y_0(x_0; a) = D'^{-1}\left(-\frac{1}{p(x_0)+1-p(a)}\right) > D'^{-1}\left(-\frac{1}{p(x_0)}\right) = y(x_0)$, since $D'' > 0$ and $p(x_0) + 1 - p(a) > p(x_0)$.

Turning to accident-prevention, recall that $C_{0;a}$ correctly perceives the marginal value of effort above the level exhibited by DM ($x > a$), but misperceives the accident to be inevitable for any level of effort below DM's choice of effort ($x \leq a$). Given this observation, it is clear that $C_{0;a}$ may only perceive as optimal two possible levels of of effort: zero accident prevention, or efficient accident prevention $x(y_0)$. To determine which one obtains, I study the sign of the cost savings $S(a, y_0)$ from choosing $x = x(y_0)$ over $x = 0$:

$$\begin{aligned} S(y_0; a) &\equiv p_0(0; a) D(y_0) - q_x x(y_0) - p_0(x(y_0); a) D(y_0) \\ &= (1 - p_0(x(y_0); a)) D(y_0) - q_x x(y_0). \end{aligned}$$

By the envelope theorem, we have $\frac{\partial}{\partial y_0} S(y_0; a) = (1 - p_0(x(y_0); a)) D(y_0) < 0$. Hence $x(y_0)$ dominates for small values of y_0, while 0 dominates for large values of y_0. Finally, we have $\overline{y}(a) < \{y | x(y) = a\}$ since $S(a; y(a)) = -q_x a < 0$, i.e. $C_{0;a}$ sets due care at 0 when $y_0 = y(a)$. ⊠

Proposition 2.2: *Let* $\underline{x} \equiv \left\{a | \overline{y}(a) = \overline{\overline{y}}\right\} \in \left(0, x\left(\overline{\overline{y}}\right)\right)$, *where* $x\left(\overline{\overline{y}}\right) < x^*$. *When the court is partially hindsight-biased, there are two potential outcomes from implementing the negligence standard:*

- *If* $q_x x^* + y^* + p(x^*) D(y^*) \leq q_x \underline{x} + \overline{\overline{y}}$, DM *chooses levels of effort* (x^*, y^*), $C_{0;a}$ *sets due care at* $\left(0, \overline{\overline{y}}\right)$, *and* DM *is held strictly liable.*

- *If* $q_x x^* + y^* + p(x^*) D(y^*) > q_x \underline{x} + \overline{\overline{y}}$, DM *chooses levels of effort* $\left(\underline{x}, \overline{\overline{y}}\right)$, $C_{0;a}$ *sets due care at* $\left(0, \overline{\overline{y}}\right)$, *and* DM *is never held liable.*

Proof of Proposition 2.2: To simplify exposition, this proof will rely on a graphical argument. The proof relies on insights from the Technical Lemma on the perception of $C_{0;a}$ about the shape of DM's reaction curves. From Technical Lemma, it is easy to see that $y_0(x_0; a) > y(x_0)$ for all a. Furthermore, the kink in $y(x; a)$ at $\overline{y}(a)$ moves upward in the $(x-y)-plane$ as the observed level of effort a increases. This obtains easily through the total differentiation of the equality that defines $\overline{y}(a)$ ($\frac{d\overline{y}(a)}{da} = \frac{\frac{\partial p(x(\overline{y}(a)),a)}{\partial a} D(\overline{y}(a))}{[p(0,a)-p(x(\overline{y}(a));a)]D'(\overline{y}(a))} < 0$).[69]

$C_{0;a}$ sets due care at the intersection of the perceived reaction curves of DM, as long as the reaction-curve with respect to damage-mitigation is flatter than the one for accident-prevention. Applying this insight, it is easy to see graphically – using the Technical Lemma – that if DM chooses a level of accident-prevention below $\underline{x}$, $C_{0;a}$ sets due care at $x > \underline{x}$ on accident-prevention. If DM chooses a level of accident-prevention above $\underline{x}$, $C_{0;a}$ sets due care at $x = 0$ on accident-prevention.

The last thing I need to show is that $\underline{x} < x\left(\overline{\overline{y}}\right)$. Assuming that $\underline{x} \geq x\left(\overline{\overline{y}}\right)$, we obtain a contradiction since $D\left(\overline{\overline{y}}\right) = q_x x\left(\overline{\overline{y}}\right) + p_0\left(x\left(\overline{\overline{y}}\right); \underline{x}\right) D\left(\overline{\overline{y}}\right) = q_x x\left(\overline{\overline{y}}\right) + D\left(\overline{\overline{y}}\right) > D\left(\overline{\overline{y}}\right)$, where the inequality follows from the fact that $p_0(x; a) = 1$ whenever $x \leq a$. ⊠

[69] $D'(y_0(a))\, dy_0(a) = \frac{\partial p(x(y_0(a)),a)}{\partial a} D(y_0(a))\, da + p(x(y_0(a)), a)\, D'(y_0(a))\, dy_0(a)$

Lemma 2.1: *Assume that the harm from an accident is fixed at a predetermined level $D(y) = h > 0$ for all y. Denote $x^*(h)$ the socially efficient level given a level of harm h, and let $\underline{x}(h) \equiv \{a|\, h = q_x x^*(h) + p_0(x^*(h); a)\, h\} \in (0, x^*(h))$. Then the optimal level of accident-prevention perceived by a partially hindsight-biased court $C_{0;a}$ satisfies*

$$x_0(a) = \begin{cases} x^*(h) & \text{if } a < \underline{x}(h) \\ 0 & \text{if } a \geq \underline{x}(h) \end{cases},$$

where $x = a$ is the level of accident-prevention exerted by DM. *We then obtain the follow outcome from implementing the negligence standard:* DM *chooses effort $\underline{x}(h)$, $C_{0;a}$ sets due care at 0, and* DM *escapes liability.*

Proof of Lemma 2.1: I fill gaps in the Proof of Lemma 2 provided in the text of the chapter. I start by establishing that $C_{0;a}$ perceives DM to potentially take only two courses of action. The first perceived possibility is that DM chooses zero accident-prevention in order to minimize losses (Option A). This dominates any effort level $x \leq a$, given $C_{0;a}$'s correct understanding that precautionary measures are costly, combined with its incorrect perception that an accident is inevitable ($p_0(x; a) = 1$). The second perceived possibility is that DM chooses $\max\{a, x^*(h)\}$ (Option B). This dominates any effort level $x > a$, given that $C_{0;a}$ correctly understands the marginal value of effort over this range, and so wants to mimic the efficient level as closely as possible.

To translate these insights into a characterization of the optimal amount of effort perceived by $C_{0;a}$, I compare the costs of Option A and Option B. $C_{0;a}$ always perceives the cost of Option A to be h. Turning to Option B, $\max\{a, x^*(h)\}$, I discuss separately the case where $a < x^*(h)$ and $a \geq x^*(h)$. When $a < x^*(h)$, $\max\{a, x^*(h)\} = x^*(h)$, and so the cost of Option B is $q_x x^*(h) + p(x^*(h); a)\, h$. Because $q_x x^*(h) + p(x^*(h); a)\, h$ is increasing in a, Option A dominates for $a < \underline{x}(h)$, while Option B dominates for $a \in [\underline{x}(h), x^*(h))$. When $a \geq x^*(h)$, $\max\{a, x^*(h)\} = a$, and so the cost of Option B is $a + h$. Because $a + h > h$; Option A necessarily dominates for $a \geq x^*(h)$. Combining these findings, I obtain the characterization of $x_0(a)$ in the statement of Lemma 2. ⊠

References

Brown, John P. "Toward An Economic Theory of Liability." *Journal of Legal Studies* 2, no. 2 (1973): 323-349.

Christensen-Szalanski, Jay JJ, and Cynthia Fobian Willham. "The hindsight bias: A meta-analysis." *Organizational behavior and human decision processes* 48, no. 1 (1991): 147-168.

Diamond, Peter A. "Single activity accidents." *Journal of Legal Studies* (1974): 107-164.

Fischhoff, Baruch. "Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty." *Journal of Experimental Psychology: Human perception and performance* 1, no. 3 (1975): 288.

Gladwell, Malcolm. "Big and bad." *The New Yorker* 79 (2004): 28-33.

Guthrie, Chris, Jeffrey J. Rachlinski, and Andrew J. Wistrich. "Inside the judicial mind." *Cornell Law Review* 86 (2000): 777.

Jolls, Christine, Cass R. Sunstein, and Richard Thaler. "A behavioral approach to law and economics." *Stanford Law Review* (1998): 1471-1550.

Kamin, Kim A., and Jeffrey J. Rachlinski. "Ex post $\neq$ ex ante." *Law and Human Behavior* 19, no. 1 (1995): 89-104.

Kolstad, Charles D., Thomas S. Ulen, and Gary V. Johnson. "Ex post liability for harm vs. ex ante safety regulation: substitutes or complements?" *American Economic Review* (1990): 888-901.

LaBine, Susan J., and Gary LaBine. "Determinations of negligence and the hindsight bias." *Law and Human Behavior* 20, no. 5 (1996): 501.

Madarász, Kristóf. "Information projection: Model and applications." *Review of Economic Studies* 79, no. 3 (2012): 961-985.

Packer, Herbert L. "Two models of the criminal process." *University of Pennsylvania Law Review* (1964): 1-68.

Rachlinski, Jeffrey J. "A Positive Psychological Theory of Judging in Hindsight." *University of Chicago Law Review* 65 (1998).

Rachlinski, Jeffrey J., Chris Guthrie, and Andrew J. Wistrich. "Probable Cause, Probability, and Hindsight." *Journal of Empirical Legal Studies* 8, no. s1 (2011): 72-98.

Schuett, Florian, and Alexander K. Wagner. "Hindsight-biased evaluation of political decision makers." *Journal of Public Economics* 95, no. 11 (2011): 1621-1634.

Shavell, Steven. "Strict liability versus negligence." *Journal of Legal Studies* (1980): 1-25.

Shavell, Steven. "A model of the optimal use of liability and safety regulation." *Rand Journal of Economics* 15, no. 2 (1984): 271-280.

Shavell, Steven. *Economic Analysis of Accident Law.* Harvard University Press, 1987.

Shavell, Steven. "Liability for accidents." *Handbook of law and economics* 1 (2007): 139-182.

Viscusi, W. Kip. "How do Judges think about risk?." *American Law and Economics Review* 1, no. 1 (1999): 26-62.

Wistrich, Andrew J., Chris Guthrie, and Jeffrey J. Rachlinski. "Can judges ignore inadmissible information? The difficulty of deliberately disregarding." *University of Pennsylvania Law Review* (2005): 1251-1345

Chapter 3
The Inter-generational and Social Transmission of Cultural Traits: Theory and Evidence from Smoking Behavior

3.1 Introduction

Cultural traits and norms, like risk-preferences, corruption, and altruism, are important in shaping individual economic behavior. The process by which such traits get transmitted from one generation to the next determines whether they survive and and how they are geographically distributed. In this study, we explain how cultures first emerge in society and how they may persist across generations even when parental preferences change over time. Our study adds to a growing economics literature that looks at the joint role of parental and social channels of cultural transmission.

Researchers from a wide range of scientific disciplines have long debated the nature/nurture question to evaluate the relative contribution of cultural (or environmental) and genetic effects on cognitive and psychological traits (Sacerdote 2011). Economists have also contributed to this debate, and have provided elaborate theory predicting that the genetic transmission of traits determines behavior (Robson and Samuelson 2011). However, this literature generally concurs with the standard Darwinian prediction of the survival of the fittest, and is at odds with evidence for the resilience of minoritarian ethnic and religious traits, e.g. among the US immigrant population. The 'melting pot' hypothesis, which suggests that characteristics and traits of the immigrant population in the US will converge over time until they resemble those of the general population, is not consistent with observed patterns. The observed behavior and characteristics of migrants conforms more closely to the 'salad bowl' hypothesis, which suggests that different ethnic and minoritarian communities can simultaneously co-exist and maintain their respective cultural identities. In fact, norms in the country of immigrant origin are found to significantly predict the behavior of second- and third-generation US immigrants (e.g Borjas 1992, Fernandez and Fogli 2009, Algan and Cahuc 2010). Starting with Bisin and Verdier (2001), the cultural transmission literature emerged to provide a rationale for this previously unexplained persistence of cultural traits (see Bisin and Verdier 2011 for a review).

Bisin and Verdier (2001) developed a model in which children acquire traits either via societal role models or via parental socialization efforts. They assume that children are more likely to acquire a given trait the more prevalent that trait is in society, and that all parents want to transmit their own cultural trait to their children. Given these assumptions, their model identifies a key feature of the social environment that ensures the co-existence of cultural traits in equilibrium. This property, which Bisin and Verdier term cultural 'substitutability', states that parents socialize their children less when their cultural trait is more prevalent in the population. Although ground-breaking, Bisin and Verdier's model fails to explain the persistence of traits that all parents want to avoid passing on to their children, even if they themselves possess them (e.g. insincerity, preferences for unprotected sex, substance use, low educational attainment). Saez-Marti and Sjogren (2008) address this failure by imposing that societal role models who belong to the minority group have a disproportionate influence on children. This condition ensures the resilience of traits that parents do not actively transmit.

Still, existing theory fails to account for two important aspects of cultural transmission. First, it says little about how traits emerge. If societal role models of a new cultural trait do not exist then it is unclear how that culture would arise in a society where all parents socialize their children to the same preferred culture. Second, the theory does not inform us about how and why parents with the same cultural trait may prefer different traits for their children. Most efforts have focused on developing a framework for contexts where parents' cultural attitudes are fixed. Yet, history is alive with examples of newly born cultural traits (e.g. the culture of social networking) and traits that switched from being perceived as virtuous or socially acceptable to being perceived as bad or undesirable, and vice versa (e.g. preferences about smoking, polluting the environment, pre-marital sex, polygyny, divorce, womens' rights etc.). Using the smoking habit as a case-study, we adapt the cultural transmission framework to account for this general class of traits.

Smoking of tobacco was first adopted before the 15th century by native Americans, who used it for recreation, medicinal purposes, or as a hallucinogenic in rituals. When Christopher Columbus and his crew discovered the American continent, they also discovered smoking of tobacco and introduced the practice to Europe. In 1854, Philip Morris made the first hand-rolled cigarettes in London's Bond Street. The second industrial revolution then saw the invention of the cigarette rolling machine, which made it possible to mass produce cigarettes cheaply. From that time and till the 1960s, the habit of smoking diffused rapidly, producing high profits for the tobacco industry. By the time of the world wars, smoking had become so socially acceptable that governments in most countries distributed cigarettes to troops as part of their regular daily rations. Some even continued to subsidize cigarette consumption during peacetime. The social perception of smoking started to change only after the publication of the Royal College of Physicians 1962 report on Smoking and Health (RCPL, 1962) and the US Surgeon General's Report on Smoking (USDHEW 1964). Those reports compiled and distilled for public consumption scientific evidence about the health consequences of tobacco use that had been accumulating for more than three decades. In the ensuing years, public campaigns against tobacco consumption followed. Over time, those campaigns and the ever accumulating evidence impacted the popularity of the smoking habit.

What this brief history of smoking teaches us is that the dynamics of the smoking culture have been shaped by the strategic behavior of a profit-maximizing industry, and the discovery and diffusion of scientific evidence. We take this lesson to theory and build the first model of cultural transmission that is able to predict both the emergence and the long-term persistence of culture in a world where outside forces may affect parents' preferred trait for their children. In our model, the flexibility in parental preferences is due to the availability and spread of information affecting perceptions about the health cost of smoking. This makes our setup more general than existing models, each of which assumes a particular distribution of parental preferences, e.g. that all parents promote their own trait (Bisin and Verdier 2001) or that all parents promote the same trait (Saez-Marti and Sjogren 2008). We also relax a standard assumption in the literature that children are more likely to acquire a given trait the more popular that trait is in society. Instead, we assume the existence of a tobacco industry which can manipulate youth smoking behavior (e.g. via advertising). We respectively define as cultural conformity and cultural distinction the positive and the negative relationship between the probability of adopting a trait and the prevalence of that trait in society. Our theory formalizes the relationship of parental and social influences that existing models predict by showing that cultural substitution in the parental channel of transmission is always tied to cultural conformity in the social channel of transmission. It also makes the novel prediction that cultural distinction is always tied to cultural complementarity. To establish support for our predictions, we carry out an empirical investigation of smoking behavior using U.S. data.

Our empirical exercise adds to a small but growing group of 'structural socialization studies' that explicitly test the properties of the transmission mechanism (e.g. Jellal and Wolff 2002, Namoro and Roushdy 2008, Patacchini and Zenou 2011, Dohmen et al. 2012). Consistent with our theoretical set-up, we model smoking participation by children as a function of parental socialization efforts and societal influences. We then estimate this function using a novel dataset. Specifically, we use data on the smoking behavior of parents and children, and on parental efforts to socialize children against substance use, from the Panel Study of Income Dynamics. We combine these data with newly constructed smoking prevalence rates from the Current Population Survey, and newly collected data on individual exposure to anti-smoking information based on the content of magazine articles. Our identification of the causal effect of parental socialization relies on state- and time-variation in parental exposure to anti-smoking information. Our identication of the causal effect of the societal influences relies on the state-level measurement of the smoking prevalence, which rules out bias due to peer-choice and residential selection, and on the use of fixed effects, which account for unobserved factors that drive both individual and group behavior. We find evidence that supports our variant of the cultural transmission theory. Using our results, we project smoking participation rates of children to future generations and show that they converge to a steady-state in which smoking persists.

The chapter is structured as follows. In Section 3.2, we summarize the health economics literature on smoking behavior and explain how this falls short of describing smoking dynamics across generations. In Section 3.3 we formally present our model. In Sections 3.4-3.6, we describe our empirical strategy, the data we use, and our empirical results. Finally, Section 3.7 concludes the chapter.

3.2 Health economics literature on the transmission of smoking

Empirically, children whose parents smoke are themselves more likely to smoke. But researchers have not established that this simple correlation reflects a causal relationship. Available evidence is often suspect because studies use poor data, small samples, or fail to account for the endogeneity of parental smoking decisions. Even the better studies have produced mixed evidence. Loureiro et al. (2010) find correlations consistent with sex-specific transmission using a UK sample. Göhlmann et al. (2010) find no sex differences using German data. Using the same data but a different identification strategy, Lillard (2011) finds that parental smoking behavior does not influence whether children start smoking (for a review of earlier studies see Avenevoli and Ries Merikangas 2003). One explanation for these mixed findings is that parental behavior may have countervailing influences on their children smoking habits and these influences may cancel out each other's effect. For example, a child of a smoker may be discouraged from smoking by parental advice and anti-smoking rules in the home (Powell and Chaloupka, 2005). Conversely, a child of a non-smoker may be attracted to smoking as an act of revolution against parental control (Huver et al., 2007). Such diverse influences of parenting styles have been overlooked by most of the empirical studies.

Further, social transmission mechanisms of smoking may also be at play. Children may copy the smoking behavior of their peers or societal role-models irrespective of their parents' behavior. This channel of smoking transmission has been the subject of a large and growing literature in economics, which consents that 'peer-effects' are important drivers of smoking participation. This general conclusion is qualitatively robust across different definitions of peer groups, e.g. schoolmates, classmates or friends (Gaviria and Raphael 2001, Powel, Tauras and Ross 2005, Lundborg 2006, Clark and Loheac 2007, Fletcher 2010); siblings (Harris and Lopez-Valcarcel 2008); partners (Clark and Etile 2006); and the population residing in the same state or prefecture (DeCicca et al 2008, Yamamura 2011). Christopoulou and Lillard (2013) argue that, in fact, societal influences on smoking participation stretch beyond current generations and across national borders. They show that the smoking prevalence of the children of British immigrants in Australia, South Africa, and the US, varies systematically with the smoking behavior of their parents' birth cohort in the UK when that cohort was at the same point in their life-cycle. The implication is that immigrant parents who grew up in a culture that tolerated (or even condoned) smoking will carry and transmit those values (consciously or subconsciously) to their children, thereby increasing the probability that they smoke.

The available theory on the economics of smoking does not reflect the intuitive processes that the empirical evidence describes. To analyze smoking behavior, researchers typically use either a myopic addiction model or a rational addiction model. These models aim to explain the life-course dynamics of the consumption of addictive goods assuming that individuals introspectively change their own preferences. The former model assumes that current consumption depends on one's own past consumption (Pollak 1970, 1976a); the latter model assumes that current consumption depends both on its past levels and on expectations of future consumption (Stigler and Becker 1977; Becker and Murphy 1988). Both models ignore the possibility that one's consumption may depend on the (past, present, or future) consumption of another person. As importantly, both models stipulate that the initial consumption of the addictive good (i.e. when previous consumption has been consistently zero) depends only on current factors and characteristics (e.g. prices). Thus, neither model is able to account for the fact that two individuals with similar characteristics who face exactly the same environment may take different smoking initiation decisions depending on their familial experiences during childhood. In effect, the models fail to achieve their purpose of describing consumption dynamics over the life-course, and they completely ignore consumption dynamics across generations. The same holds for subsequent models that were built to reconcile the myopic and rational addiction theories (e.g. Orphanides and Zervos 1995, 1998).

The economic theory on social interactions, which was developing concurrently with the theory of addiction, offers a less restrictive framework of analysis. The benchmark study by Pollak (1976b) developed a model of habit formation and learning, allowing individual preferences to depend on others' behavior, which either provides information about the costs and benefits of behavior, or establishes a point of reference. However, rarely have the theoretical implications of this approach been drawn with respect to smoking behavior, and in the few occasions that they have, they have focused on peer effects and have ignored the parental channel of transmission (e.g. Nakajima 2007, Poutvaara and Siemers 2008). In this chapter, we draw on the literature of intergenerational transmission of cultural traits to extend the theoretical analysis in this direction.

3.3 The model

3.3.1 Environment

Assume that there are an infinite number of periods ($t = 1, 2, 3, ...$), and each individual is alive for two consecutive periods. A person born in period t is a child in period t and an adult in period $t+1$. Each individual bears one child in adulthood. Hence, in period $t+1$ the population consists of two overlapping generations: adults (born in period t) and children (born in period $t+1$). We use "he" to refer to a generic child and "she" to refer to a generic adult.

Individuals are either smokers or non-smokers; we call q_t the proportion of the youth who smoke in period t. We then denote Q_{t+1} the proportion of adult smokers in period $t+1$. An individual's smoking behavior need not be constant throughout the course of a lifetime, since a smoker may decide to quit smoking. To simplify the exposition, we however impose $Q_{t+1} = q_t$. This is consistent with the literature on cultural transmission, which assumes a cultural trait is acquired once and for all during childhood.[70] Our results are qualitatively robust to extending the model to one where parents may quit smoking in adulthood.

3.3.2 Socialization process

Children are born without predefined traits and acquire their smoking behavior through a transmission process. A child is first exposed to his parent's influence, a process we refer to as **vertical transmission**. The parent's *preferred* trait need not coincide with her *exhibited* trait, and so our specification does not force parents to socialize their children to their own traits. In particular, a smoker parent may choose to socialize her child away from smoking, while remaining a smoker herself. The driving force for this divergence between own behavior and desired behavior could be, for instance, that quitting has addiction costs that are only incurred by someone who already smokes. To our knowledge, Saez-Marti and Sjogren (2008) and Patacchini and Zenou (2011) are the only studies to have theoretically investigated in a cultural transmission framework the possibility that parents agree on the trait to be passed on to future generations (e.g. educational attainment).

[70] By implicitly assuming that socialization is a function of a parent's youth (rather than adult) smoking behavior, we are imposing that attitudes towards smoking are acquired early on in life. This modeling assumption allows us to abstract away from the strategic considerations that would arise from a situation where it would be Pareto optimal for all parents to quit smoking for their childrens' sake, but who cannot credibly commit to doing so because they each have an incentive to freeride on others' efforts.

In contrast, the bulk of the literature on cultural transmission has been motivated by the study of traits like religion, where parents want to promote their own trait.

Bisin and Verdier call 'direct' transmission the process of having parents socialize children to their own traits. 'Vertical transmission' can thus be thought of as a generalization of 'direct transmission'.[71] We will then refer to a parent socializing her child to her own trait as *direct socialization*, and to a parent socializing her child to a trait other than her own as *active socialization*. From here on, we focus on the case where all parents want to discourage their children from becoming smokers. A smoker parent does not directly socialize her child to smoke, nor does a non-smoker parent actively socialize her child to smoke. This feature arises endogenously once we let parents choose the extent of vertical transmission, as it would be counterproductive to ever encourage children to acquire the unwanted cultural trait. In the terminology of the model, non-smoker parents directly socialize their children to their own trait with probability $d(q_t)$, while smoker-parents actively socialize their children to the other trait with probability $a(q_t)$. With remaining probability $1-d(q_t)$ for non-smoker parents, and $1-a(q_t)$ for smoker parents, the child acquires his trait through non parental influences, a channel which the literature typically labels as **horizontal transmission**. With probability $S(q_t)$ / $1-S(q_t)$, the child then becomes a smoker/non-smoker. We think of $S(q_t)$ as reflecting not only the influence of societal role models but also the advertising efforts of a profit-maximizing tobacco industry, as well as the child's own choices.

Throughout the analysis, we impose one of two possible assumptions on the vertical transmission process: *cultural substitution* or *cultural complementarity*. The concept of cultural substitution has been a cornerstone of the literature starting with Bisin and Verdier (2001), while cultural complementarity has received less attention (e.g. Bisin, Topa and Verdier 2004; Patacchini and Zenou 2011).

Assumption V1 (Cultural Substitution)**:** *$d(q_t)$ and $a(q_t)$ are increasing in q_t, and $d(0) = a(0) = 0$.*

Assumption V2 (Cultural Complementarity)**:** *$d(q_t)$ and $a(q_t)$ are decreasing in q_t, and $d(1) = a(1) = 0$.*

[71] Note however that our terminology deviates from the literature, which interchangeably uses "vertical" and "direct" transmission to refer to children mimicking their parents.

Assumption V1 says that a parent's vertical transmission effort is an increasing function of the unwanted trait in society, and that parents exert no effort when the unwanted trait is absent from society. When smoking is the unwanted trait, parents want to make sure their children do not fall prey to the influence of smoker role models, and so $d(q_t)$ and $a(q_t)$ are increasing functions of smoking prevalence q_t. Assumption V2 says that a parent's vertical transmission effort is a decreasing function of smoking prevalence in society. This phenomenon can emanate in two possible ways. Parents may either give up on socializing their children when the outside threat becomes greater, or the threat of the unwanted trait might be decreasing with its prevalence in society.

We also impose one of two assumption on the horizontal transmission process: *cultural conformity* or *cultural distinction*. While cultural conformity has been an implicit feature of existing models on cultural transmission, cultural distinction has been largely overlooked by theoretical studies. In a recent paper, Bisin et al. (2013) call cultural conformity (distinction) the drop (increase) in psychological costs of interacting with a member of an outside cultural group when that outside group becomes more dominant. They use those notions to understand the implications of ethnic identify on marriage outcomes in a cultural transmission framework. We, however, define those concepts differently:

Assumption H1 (Cultural conformity): $S(q_t)$ is increasing in q_t.

Assumption H2 (Cultural distinction): $S(q_t)$ is decreasing in q_t.

A large literature on identity formation in psychology, sociology, and political science argues that cultural or ethnic minority groups may either pursue cultural assimilation into the majority population or they may strive to keep their distinct identities.[72] We follow this reasoning to assume that the child culturally distinguishes himself when he horizontally adopts the minority culture, and he culturally conforms when he horizontally adopts the majority culture. Thus, cultural distinction predicts that the greater smoking prevalence is in society, the less likely the child is to adopt this trait. Cultural conformity makes the opposite prediction, that the greater smoking prevalence is in society, the more likely the child is to adopt this trait.

As we noted above, our definition of cultural conformity has been implicitly assumed in the literature as part of the horizontal socialization. Existing work has thought of $S(q_t)$ as representing

[72] See Bisin and Verdier (2011) for a survey of this literature and a list of references.

a *matching process*, i.e. the likelihood a child is matched to a smoker role model from the population of adults (random-matching (e.g. Bisin and Verdier 2001; Bisin et al.2013); non-random matching (e.g. Bisin et al. 2004; Saez-Marti and Sjogren 2008)).[73] This matching process must satisfy two basic properties: (i) A child can only be matched to a non-smoker/smoker when there are only non-smokers/smokers ($S(0) = 0$ and $S(1) = 1$), and (ii) the likelihood of being matched to a smoker increases with the proportion of smokers in society (i.e. our definition of cultural conformity). As we will show in the empirical section of this chapter, there is strong support for the possibility of having cultural distinction, which cannot be accommodated by existing models. More importantly, the study of dynamics in Section 3.3.5 reveals that under cultural substitution, assuming $S(0) = 0$ leads to the disappearing of the smoking culture in steady state, which is at odds with real-world behavior. The mechanism for horizontal transmission that we propose addresses this gap in the theoretical literature.

3.3.3 Endogenous vertical transmission

We now explicitly define the likelihood that a child follows his parent's desire to not smoke as a function of parental socialization. We find the optimal level of parental investment, and derive comparative statics for the effect of smoking prevalence and health costs of smoking.

As before, the subscript $i \in \{0, 1\}$ denotes the adult's smoking behavior. When the child is born, his parent decides how much to invest in his anti-smoking socialization, $\lambda_i \geq 0$. The cost of socialization is $c(\lambda_i)$, and causes the child to become a non-smoker with probability $v(\lambda_i)$. With remaining probability $1 - v(\lambda_i)$, the child becomes a smoker with probability $S(q)$, and a non-smoker with probability $1 - S(q)$.

We normalize a parent's utility from not seeing her child smoke to 0. We then denote by $u_i(H) < 0$ the utility of having a child who smokes, where H represents the *perceived* (but not necessarily actual, $\overline{H}$) detrimental health effects of smoking. All parents alike value their children's health status, and so $u_i(H)$ decreases with the perceived health costs of smoking. Formally, an

[73] Saez-Marti and Sjogren (2008) refer to as conformism a non-random matching process that favors the dominant trait. Their definition of conformism thus differs from both our and Bisin et al. (2013)'s definition of cultural conformity.

adult's utility function can be written as follows:

$$U(\lambda_i; q, H) \equiv (1 - v(\lambda_i)) S(q) u_i(H) - c(\lambda_i),$$

where a parent's investment is associated with a probability $(1 - v(\lambda_i)) S(q)$ of seeing her child smoke. Proposition 3.1 gives the optimal interior investment in the anti-smoking acculturation of a child.

Proposition 3.1*: Let $v(.)$ be a differentiable, increasing and concave function, and $c(.)$ be a differentiable, increasing and convex function. When interior, a parent's optimal investment choice λ_i^* solves*

$$\frac{v'(\lambda_i^*)}{c'(\lambda_i^*)} = \frac{1}{-u_i(H) S(q)}. \tag{9}$$

Proof of Proposition 3.1: See Proofs Section.

The assumptions on $v(.)$ and $c(.)$ are standard to ensure the existence of a unique solution to the maximization problem. They state that (i) more effort leads to a greater likelihood (cost) of dissuading the child from acquiring the unwanted behavior, and (ii) the effectiveness (cost) of this investment decreases (increases) in the amount of effort that is invested. The optimal investment choice simply equates the marginal cost of investing to its expected marginal benefit. The assumption that $u_i(H) < 0$ for $i = 0, 1$ makes it not optimal for any parent to invest in pro-smoking culturalization, as this would promote the parent's undesired trait. Had we assumed instead that smoker parents perceive a non-negative net utility from smoking, $u_1(H) \geq 0$, we would be in the world of Bisin and Verdier (2001) where each type of parent promotes her own trait. Ceteris paribus, this possibility is more likely to arise when the health costs of smoking are small, since both parents perceive children's utility to be decreasing the more harmful the smoking habit is. From the result in Proposition 3.1, we can then derive comparative statics on the investment choice:

Proposition 3.2: *Performing comparing statics on the optimal investment derived in Proposition 3.1, we obtain the following predictions:*

- *Prediction 1: Parents who perceive* larger *health costs from smoking invest* more *in anti-smoking culturalization.*
- *Prediction 2: We have cultural substitution iff cultural conformity holds, and cultural complementarity iff cultural distinction holds.*

Proof of Proposition 3.2: See Proofs Section.

Prediction 1 is straightforward; it states that parents invest more in the anti-smoking culturalization of their children the more serious the health costs of this habit. Prediction 2 says that vertical socialization follows cultural substitution when horizontal socialization follows cultural conformity, which is a well-known theoretical result in the literature, starting with Bisin and Verdier (2001). It implies that when a higher smoking prevalence increases the likelihood that a child smokes, it is more worthwhile for the parent to exert socialization effort. However, Prediction 2 also says that vertical socialization follows cultural complementarity when horizontal socialization follows cultural distinction, which is a result that is not emphasized in the literature. It implies that when smoking prevalence decreases the likelihood that a child smokes, it also lowers the value of socialization to the parent. Patacchini and Zenou (2011) produce the opposite finding in a cultural transmission model augmented with peer effects at the vertical transmission stage. Their result requires assuming that both the cost and marginal cost of socialization are increasing in the proportion of smokers. Like the rest of the literature, we instead favor a framework where society only plays a role through horizontal transmission; i.e. in the event that the parent is unsuccessful at socializing her child to the desired trait.

3.3.4 Endogenous horizontal transmission

To complete the model, we explicitly define the likelihood that the child adopts smoking through non-parental channels. Before parents decide how much to socialize their children, a profit-maximizing monopolist chooses a level of investment θ into increasing the appeal of smoking to the youth. Such a feature could arise, for example, through celebrity endorsements of this habit. Advertising θ can be thought of as stimulating "demand", but to have diminishing marginal returns, so that $\frac{\partial S(q,\theta)}{\partial\theta} > 0$ and $\frac{\partial^2 S(q,\theta)}{\partial\theta^2} < 0$.[74] The function $\kappa(\theta)$ represents the convex cost of advertising ($\kappa' > 0$ and $\kappa'' > 0$). Keeping the firm's investment constant, we assume that $S(.)$ can still be thought of as a matching process, so that $\frac{\partial S(q,\theta)}{\partial q} > 0$. We finally impose a regulatory condition to ensure an interior level of investment for the firm, $\frac{\partial S(0,\theta)}{\partial\theta} > \kappa'(0)$. Proposition 3.3 describes the equilibrium of the game implied between the monopolist and parents.

Proposition 3.3: *In a subgame perfect equilibrium, we have $S(0,\theta^*(0)) > 0$, which ensures a heterogeneous distribution of traits in steady state. Moreover, both cultural substitution and cultural conformity may obtain.*

Proof of Proposition 3.3: See Proofs Section.

$S(0,\theta^*(0)) > 0$ follows directly from assuming an interior solution to the problem. It ensures that the tobacco company would always want to invest a positive amount in making the smoking habit emerge. The second part of the result comes from a study of the sign of $\left.\frac{dS}{dq}\right|_{\theta^*} = \frac{\partial S}{\partial q} + \left.\frac{\partial S}{\partial\theta}\right|_{\theta^*}\frac{\partial\theta^*}{\partial q}$. We have cultural complementarity whenever this quantity is positive, and cultural substitution otherwise. While determining the sign of this quantity involves the interaction of numerous terms, it is possible to talk loosely about when each possibility obtains using the findings in Proposition 3.2. For example, cultural distinction obtains when: 1) an increase in smoking prevalence decreases the value of the investment sharply ($\frac{\partial^2 S(q,\theta)}{\partial\theta\partial q}$ negative enough), or 2) smoker parents are less tolerant of their children becoming smokers ($a > d$). The main take away message is that both cultural conformity and cultural distinction are realistic possibilities. We leave it to the data to tell us when each obtains.

3.3.5 Dynamics

We summarize the period $t+1$ transmission process into a transition matrix $P_t \equiv \begin{bmatrix} P_{t0} \\ P_{t1} \end{bmatrix}$, where $P_{t0} \equiv (1-d(q_t))\,S(q_t)$, $(P_{t1} \equiv (1-a(q_t))\,S(q_t))$ gives the proportion of children of non-smokers (smokers) who adopt smoking. This matrix is subscripted by the time period t since transmission

[74] We put the word "demand" in quotation marks given that children do not make any active choice in the cultural transmission framework.

is a function of the parents' smoking behavior, which is acquired in their youth. We can represent the evolution of smoking in society through the equation

$$\dot{q} = P_t^T \begin{bmatrix} 1 - q_t \\ q_t \end{bmatrix} - q_t, \tag{10}$$

where the superscript T indicates the transpose of a matrix. The following result describes the steady state behavior of the system.

Proposition 3.4: *In a steady state, there always exists a fraction of non-smokers. Moreover, under cultural complementarity the smoking habit always persists, while under cultural substitution it persists as long as* $S(0) > 0$.

Proof of Proposition 3.4: See Proofs Section.

Under cultural complementarity, we must have cultural distinction (Proposition 3.2), and therefore children always have the proclivity to reject the status-quo so that neither traits disappear in steady state. Under cultural substitution, smoking never becomes the unique trait since parents have an incentive to prevent this from happening. In contrast, smoking may disappear if no outside factor forces it to persist. Given Proposition 3.3, we know that the existence of a tobacco industry can guarantee the coexistence of both traits under this scenario.

3.4 Empirical strategy

We use the following baseline specification:

$$\begin{aligned} Pr(ever\ smoke = 1)_c &= \alpha_0 \\ &+ \alpha_1 * Pr(socialization = 1)_p \\ &+ \alpha_2 * sm.prevalence\ of\ role\ model\ population_{cs} \\ &+ \sum_j \alpha_{3j} * X_{jcs} \\ &+ \nu_c \end{aligned} \tag{11}$$

$$\begin{aligned} Pr(socialization = 1)_p &= \beta_0 \\ &+ \beta_1 * exposure\ to\ health\ information_p \\ &+ \beta_2 * sm.prevalence\ of\ role\ model\ population_{cs} \\ &+ \sum_j \beta_{3j} * X_{jcs} \\ &+ \upsilon_c \end{aligned} \tag{12}$$

Equation (3) is the empirical counterpart of P_T, as described in the previous section. It is a structural form equation that treats $Pr(socialization = 1)$ as the endogenous regressor. Equation (4) is the empirical counterpart of (1). In this first-stage equation we identify parental socialization using different indicators of the parent's *exposure to health information* as instruments. X denotes exogenous control variables; α denotes a structural parameter; and β denotes a reduced-form parameter. Indexes c and p stand for child and parent, respectively; s stands for state; and j identifies each characteristic (individual, parental, or state) that we include as a control variable. Finally, ν and υ are the jointly distributed error terms.

We estimate (3) and (4) as a system by IV probit, even though this method is meant to be used when the endogenous regressor is continuous rather than binary. Because Heckman's (1978) maximum likelihood bivariate probit was built to accommodate binary endogenous regressors, it would have been more appropriate to use in our case. However, we choose not to use it because it is computationally cumbersome,[75] and it does not significantly outperform IV probit or even IV linear

[75] Researchers find that to run a bivariate probit often takes 10 or 20 times as long as other similar models, and that standard statistical software like Stata and R frequently fail to find the maximum of the likelihood (e.g. Freedman and Sekhon 2010)

probability models in terms of accuracy (see Nichols 2011 and references therein). To confirm the latter point, we test the robustness of our baseline specification to a range of alternative estimation methods, including the bivariate probit.

To statistically identify exogenous variation in parental socialization we assume that, controlling for the child's own exposure to anti-smoking articles, we can exclude parental exposure to health information as a direct determinant of the child's decision to smoke. To test this exclusion restriction, we calculate the Amemiya-Lee-Newey (ALN) minimum χ^2 statistic under the null that the instruments are valid (i.e. uncorrelated with the error term) and correctly excluded from the outcome equation.[76] To test whether our instruments have weak explanatory power, we calculate the χ^2 statistic under the null that the instruments are jointly statistically insignificant in the reduced form. We also calculate the Hausman χ^2 statistic to test whether there is a statistically significant difference between IV probit and probit (naive) estimates. The null of this test is that the probit model provides both consistent and efficient estimates while IV probit estimates are only consistent, and that the difference between the two is normally distributed with mean zero. Further, we calculate the Wald χ^2 statistic to test the null that the correlation coefficient between ν and υ is zero and, therefore, $Pr(socialization = 1)$ can be treated as exogenous. Finally, we check the robustness of our baseline estimates to the inclusion of a wide range of controls and instruments.

To statistically identify the social transmission of the smoking trait we rely on the fact that the smoking prevalence of the role-model population is measured at the state-level. Because we can plausibly assume that state-specific smoking prevalence is exogenous to the parental choice of the state of residence and it cannot be affected by endogenous peer-choice, we rule out selection and simultaneity bias from the estimated effects. Bias due to exogenous correlated effects, however, remains a possibility (Manski 1993, 2000). The smoking prevalence of the role-model population is the aggregation of individual behavior which (depending on how the role-model population is defined) may include the parent or the child. Thus, our estimate of α_2 may reflect the fact that individuals in a given state have similar smoking behavior because they have unobserved similar characteristics or because they are exposed to the same institutional or contextual factors ('Manski's reflection problem'). To account for such unobserved common factors, we follow the health economics litera-

[76]The Amemiya-Lee-Newey test is only possible after running the two-step Newey (1985) IV probit estimator. All other results we present in this chapter are derived using the maximum likelihood IVprobit estimator.

ture and use a fixed effects specification (see, for example, Nakajima 2007 and references therein). Because, as we describe below, the smoking prevalence of the role model population varies by state and child age, we include a full set of state and age fixed-effects. We thus identify causality of the social effects by using variations in the proportion of smokers between age-groups within a state.

We use the results to assess whether socialization by parents and role models affect a child's smoking decision, to identify the relative contribution of the two types of socialization to the transmission of the smoking culture, and to test important properties of the transmission process; namely, cultural substitution versus complementarity, and cultural conformity versus cultural distinction. Finally, we use the structural parameter estimates to forward project how the rate of ever-smoking of 10-18 year olds will evolve under different policy scenaria.

3.5 Data

Our empirical analysis exploits new as well as existing data in novel ways. Below we discuss the source and construction of each type of data. Table 3.1 provides summary statistics of selected variables.

3.5.1 Individual level data on children and care-givers

We draw individual level data on children aged 10-18 from the 2002 and 2007 waves of the Panel Study of Income Dynamics (PSID) - Child Development Supplement/Transition to Adulthood (CDS-TA) surveys, and supplementary data on parents of each child from the main family files of the PSID. The CDS-TA sample was originally drawn from PSID families with children 0-12 years in 1997 and reinterviewed in 2002 and 2007. We use sample weights provided by the PSID to account for unequal probabilities of being selected into the CDS sample and for differential attrition rates of PSID and CDS participants. The 2002 and 2007 wave of CDS-TA directly surveys children age 10 and older about tobacco use and other types of behaviors. These data were collected by an audio computer assisted self-interview. In such interviews, youth listen to the questions through a headset and record their responses directly into a laptop computer. Neither the child's parents nor the interviewer knew how s/he answered the questions. Aquilino (1994) documents that this method generates more accurate data on socially sensitive topics such as psychological well-being, sexual behaviors, and experiences with tobacco, alcohol, and drug use. We use the data on whether a child ever smoked (defined on the survey as smoking at least 1 cigarette every day for 30 days) to represent $Pr(ever\ smoke = 1)_c$. Eighteen percent of our sample smoked at some time between age 10 and 18.

The CDS-TA surveys asks questions not only of the child but also of the person in the PSID household who identified herself/himself as the 'primary' care-giver (PCG) of that child. Table 3.1 documents that biological mothers comprise 93 percent of self-identified care-givers, six percent of care-givers are biological fathers, and the rest are adoptive mothers or step-mothers. Because of the disproportionate share of biological mothers in our sample, we cannot separately model how cultural transmission varies with the nature of the parent-child relationship.

While the CDS-TA surveys do not ask questions that are specifically about smoking socialization efforts of the PCG, the surveys do collect information that is likely to proxy for it. Ideally the surveys would ask each PCG to report how much effort she spends socializing her child about the health risks of smoking. Instead the CDS/TA surveys asked each care-giver to report how frequently during the past month she talked to each child about the dangers of substance use (e.g. drinking alcohol or taking drugs). The survey specified five response categories that ranged from "not in the past month" to "every day." While these data are not ideal, we expect answers to them to be correlated with the conceptual variable of interest. In addition, we identify variation in the pattern of responses in these data using variation in information that is specifically about health risks of smoking. As a result, though imperfect, the CDS data are likely to proxy well for the conceptual variable of interest. Table 3.1 shows that 21 percent of care-givers reported they had not spoken with their child about the dangers of substance use in the past month. Most (42 percent) care-givers discussed this subject once or twice per month. However, 37 percent of care-givers discussed substance use at least once a week during the previous month.

In addition to the socialization data, we draw CDS-TA data on age, sex, race, and religion of the child; household income; family size; and measures of parenting styles, reading habits, and employment status of the PCG.

We also draw data on the smoking behavior and educational attainment of the PCGs from the main PSID files. We use data on smoking behavior from the 1986, 1999, 2001, 2003, 2005, 2007, and 2009 family files and data from a special 1990 questionnaire administered to all PSID household members over the age of 55 to construct our measure of whether a parent ever smoked. Finally, we draw data on years of completed schooling from all waves of the PSID. In our sample, 45 percent of PCGs smoked at some point in their lives and the average PCG has completed 13 years of schooling.

3.5.2 Smoking prevalence rates of the role model population

To construct measures of role-model smoking behavior, we use data from the Tobacco Use Supplements to the Current Population Survey (TUS-CPS). Sponsored by the National Cancer Institute and administered as part of the U.S. Census Bureau's continuing labor force survey, the TUS-CPS data have been collected intermittently since 1955. We use responses from 21 surveys conducted in August 1967, August 1968, September 1989, September 1992, January and May 1993, September 1995, January and May 1996, September 1998, January, and May 1999, June and November 2001, February 2002, February, June, and November 2003, May and August 2006, and January 2007. Each survey asks respondents: "Have you ever smoked regularly?"; "If yes, what is the age when you started?"; "Do you currently smoke?"; and "If not, what is the age when you last smoked regularly?".

After dropping multiple observations for each individual across monthly waves of the same calendar year, we pool all data from these waves and use the smoking questions to construct the smoking history of every TUS-CPS respondent. To do this we identify all respondents who ever smoked and who report a start age, a current smoking status, and a quit age (former smokers only). We then assume that a person smoked in every year between the age she started and either the age at the survey date (current smokers) or the age she quit (former smokers). Because in each calendar year our sample includes all respondents who were alive in that year and retrospectively answered the smoking questions in any later year, we start with an enormous sample of current, ever, and never smokers (approximately 81 million observations). We combine our computed smoking life-histories with data on the state of residence (at the time of the survey) to construct smoking prevalence rates by sex, cohort, state, and calendar year (weighted by the CPS sampling weights).

To match our empirical specification as closely as possible with our theory, we assume that a child's role model is drawn from his parent's generation. This assumption means that we create smoking prevalence rates of the role model population using males and females who are 20-29 years older than the child.[77] Table 3.1 shows that among the potential social role models of children in

[77] i.e. we assume an average generation gap of 25 years, which is slightly smaller than the average generation gap that we

our sample, the average smoking prevalence rate is 26 percent. Later in the analysis we experiment with different definitions of the role-model population (by age and sex).

To illustrate the variation in the data we develop, Figure 3.1 plots the smoking rates by sex, state, and calendar year for the social role models of children who are age 14 at the time of the survey (i.e. at mean age). While these data have rich variation across all dimensions, the variation we can exploit is limited because we only observe a child's smoking behavior and his parent's socialization effort in 2002 and 2007. Despite that limitation, plenty of variation remains available to us: across states, gender, and by child age. Figure 3.2 showcases the data that we actually use in the analysis.[78] Clearly, the smoking rates vary significantly by gender and state, but note that, because the state-specific curves cross, the rates also vary by age of the child/parent generation.

3.5.3 Information about the health risks of smoking

To instrument parental socialization we use temporal and geographic variation in exposure to information about the health risks of smoking. The basic data consist of counts of articles published between 1924 and 2009 that warn readers about the health risks of smoking. We use counts of articles published in each of more than 21 popular consumer magazines. To generate additional geographical, but also temporal variation, we also exploit data on the number of issues of each magazine that were sold in each state in each year.

The data on articles were generated by first searching two electronic databases (ProQuest and the Historical Reader's Guide to Periodical Literature) using a keyword search on "smok* and cancer," "smok* and health," "cig* and cancer," "cig* and health" and similar text strings. Successive searches produced roughly 5,000 titles of articles published between 1890 and 2009. Two undergraduate research assistants then independently reviewed all 5,000 titles to identify articles that potentially warned about the health risks of smoking. This review eliminated roughly 2,500 articles that focused on the effect of the health risk information on financial returns of tobacco companies, tobacco growing agriculture, and international trade in tobacco. The remaining set included articles whose

observe in our data (28 years).

[78]Note that instead of 2007 data we use data that correspond to 2006 because the 2007 survey was conducted in January. Aside from the fact that the sample is smaller, smoking behavior in January does not represent average smoking behavior throughout the year because more people quit smoking in January to implement a New Year's resolution. Many of these people fail in their resolution. Consequently, smoking prevalence rates in January are lower and more widely distributed than smoking prevalence rates over the whole year.

titles suggested that the articles discussed content about risks individuals faced. A team of research assistants collected copies of all articles and read them. Two of them independently rated the articles as a) "pro-smoking," b) "neutral," and c) "anti-smoking" when they judged that an article conveyed to readers the impression that smoking a) improved, b) did not affect, or c) degraded the health of smokers. Any disagreement was discussed and resolved. The resulting list of articles generated a list of magazines in which the articles appeared.

We then compiled data on sales of each of those magazines in every state in each year. We got the sales data from the Audit Bureau of Circulation. The Audit Bureau of Circulation is an organization that publishers voluntarily join. Its sole purpose is to audit and verify circulation figures the publishers provide to them. Their independent auditing provides a valuable service to publishers because they charge advertisers more for space in more widely circulated magazines. Advertisers therefore demand (and publishers willingly provide) an independently verified count of circulation. The magazine circulation data vary by month, year, and state.

We assume that, when a magazine is sold, it is seen by all members of the household in which the purchaser resides.[79] To capture this exposure, we divide estimates of each state's population from the Current Population Reports of the Census Bureau by 2.3 (the average household size) and divide the number of issues sold in each state in each year by that number. The resulting Figure 3.is an estimate of the fraction of each state's population that read each magazine (in each year). We then multiply the fraction of each state's population that read each magazine by the number of articles that appeared in that magazine. This step yields the exposure of a randomly drawn person from a given state to an article that appeared in a given magazine in a given year. Finally, we sum across all magazines in which an anti-smoking article appeared. The final data proxy for the total potential exposure to anti-smoking magazine articles in a given state in a given year. Currently we compute this sum using articles that appear in 21 magazines that accounted for 70 percent of all anti-smoking magazine articles produced by the above searches. Formally our measure is given by:

$$\text{Anti-smoking articles read}_{st} = \sum_{m=1}^{21} \text{Articles}_{mt} \frac{\text{Issues}_{mst}}{\text{Population}_{st}/2.3} \tag{13}$$

where s denotes state, t denotes calendar year, and m denotes each of 21 magazines.

[79] In fact we only assume a household member *potentially* sees the article. From now on we use the terms "exposure" and "potential exposure" interchangeably.

Figure 3.3 plots the resulting measure for all states between 1929 and 2009. We use these data to compute several alternative measures of exposure to anti-smoking information of both PCGs and children. The two measures we select to use in the baseline specification are: (i) the accumulated sum of articles (potentially) seen by the PCG or the child since age 10, which should capture the degree of exposure; and (ii) the standard deviation in the articles read since age 10, which should capture the *in*frequency of the information flow. Similar measures have been shown to predict changes in consumption of fats and oils as information developed and spread about the health risks of consumption of saturated, monounsaturated, and poly unsaturated fats (Chern et al. 1995). We start counting exposure from age 10, assuming that it is the earliest age a child can comfortably read. Year 1929 is the earliest year in which a PCG in our sample was age 10, and year 1994 is the earliest year in which a child was age 10. Thus, our measures of exposure for PCGs (i.e. our instruments) encompasses all temporal and state variation illustrated in Figure 3.3. In contrast, our equivalent measure for children (which we use as a control) encompasses the variation illustrated in the shaded area only. The 'typical' PCG in our sample has been exposed to about 12 anti-smoking articles since age 10, while the corresponding number for the 'typical' child is 4.

3.5.4 Indicators of the economic environment

In our regressions, we control for time-varying state-specific economic factors that may affect the child's probability to smoke or the PCG's socialization efforts. Specifically, we control for state and federal cigarette taxes and state per-capita income. We use the measure of 'full' taxes on cigarettes described in Lillard and Sfekas (2013). This measure is the sum of the state and federal cigarette tax and the per pack escrow payments that are required on the 1998 "Master Settlement Agreement" between the four major cigarette manufactures and the US states. Viscusi and Hersh (2011) and Lillard and Sfekas (2013) document that this payment is functionally equivalent to a per pack cigarette tax. We draw data on real per capita state income from the Regional Economic Information System of the Bureau of Economic Analysis in the U.S. Department of Commerce (SA05 series) and adjust them in units of real 2008 dollars. To illustrate the variation in these two variables, we plot them by state and year in Figures 3.4 and 3.5.

3.6 Results

3.6.1 Vertical versus horizontal transmission of the smoking trait

The naive way to think about the cultural transmission of the smoking trait would be to assume no reverse causality between parental socialization efforts and the childrens' decision to smoke. Table 3.2 presents estimates of equation (3) under this assumption; that is, by treating $Pr(socialization = 1)$ as exogenous. In column 1 we use the socialization data in their 'raw' form; i.e. in five categorical dummy variables: not in the past month (the reference category), once/twice a month, once a week, several times a week, and daily. In columns 2-5 we dichotomize the data into a single dummy to denote that children are socialized with 'at least' a given level of frequency (at least once/twice per month, at least once a week etc.). The resulting estimates vary across the combinations. In column 1, PCG socialization efforts are associated with a higher probability that the child ever smokes, even though that association decreases as the frequency of socialization increases. The estimates in column 2 also suggest a positive association, while columns 3-5 suggest a weak negative association (which in 3 and 5 is statistically insignificant). The probable cause of these counterintuitive results is the endogeneity of the socialization variable. Children may be less likely to smoke when their parents socialize them against it, but it is also likely that a PCG will discuss substance use more often if a child smokes or the parent suspects a child is likely to smoke. To isolate the former effect we abandon the naive approach.

Table 3.3 presents our baseline estimates of jointly determined equations (3) and (4), as described in Section 3.4.[80] To simplify the analysis, we dichotomize the socialization variable, $Pr(socialization = 1)$, to indicate the probability that the PCG socializes the child about the dangers of substance use at least once a week.[81] As we mentioned earlier, we instrument the PCG's socialization efforts with the parental exposure to smoking-related health information since age 10 and the infrequency of that

[80] From this point onwards all regressions are estimated by IV probit. See Table 3.4 for a set of robustness test of the baseline estimates to alternative methods of estimation.

[81] We dichotomize the socialization variable because the coefficients on the different versions of the socialization variable presented in columns 2-5 of Table 3.2 suggest that socialization categories "once a week", "several times per week" and "every day" produce results that are similar among them but much different to the results produced by category "once/twice a month". This implies that the response distribution we observe could be a mixture of two distributions, each capturing a different kind of decision process. Nonetheless, our baseline results show low sensitivity to the definition of the dichotomized socialization variable. See Table 3.5 for the robustness analysis. In a similar exercise, Patacchini and Zenou (2011) use the frequency that a parent reads to a child to capture parental effort to cultivate interest in education to the child. They dichotomize their socialization variable the same way.

exposure. In contrast to the results in column 3 of Table 3.2, the coefficient on the instrumented socialization variable is negative and statistically significant, suggesting that a parent's effort to socialize her child is effective. The performance of the instruments in the first-stage equation is satisfactory: they are statistically significant and the sign of their coefficients are in the direction described in Prediction 1. Those PCGs who are exposed to more anti-smoking articles on average also exert more effort to socialize their children. Holding the average level of exposure constant, PCGs socialize their children less if their exposure varies more over time. That is, PCGs socialize their children more when they are exposed to a constant stream of information about the health risks of smoking compared to PCGs who see the same number of articles on average but who see no articles in some years and many articles in others. The diagnostic test results corroborate the good performance of our instruments and of the baseline specification in general.

We should note that a factor contributing to instrument validity is that the baseline specification controls for the child's exposure to anti-smoking articles since age 10. Again, the estimated coefficient on this variable makes economic sense: a higher information exposure of the child is associated with a lower probability that the child ever smokes and with a lower socialization effort by the PCG (as s/he now relies on the external information to do the job). Although the statistical significance of these effects is somewhat weak, it is important to mention that removing the measure of child exposure from the estimation significantly impacts the ALN test of over-identification. In this case, the χ^2 statistic increases to 3.255 and the probability values drops to 0.071, so that we reject the hypothesis that the instruments are valid at the 10% level of significance.

In all regressions reported in Tables 2 and 3, the probability that the child ever smokes increases with the state-specific smoking prevalence rates of the role-model population, providing evidence for cultural conformity. In the first-stage of Table 3.3, the socialization effort of PCGs also increases with the smoking prevalence in the role model population and it is statistically unrelated to parental smoking status, providing evidence for cultural substitution. The co-existence of cultural substitution and conformity supports Prediction 2 of the model. It suggests that all parents wish to discourage smoking and, because they know that their children will conform to societal trends, they will increase their anti-smoking socialization efforts when smoking becomes more popular in society (we expand this discussion in Section 3.6.3).

To compare the importance of the parental (vertical) and social (horizontal) channels of transmission of the smoking habit, we calculate marginal effects of 'equivalent' changes in parental socialization efforts and the smoking prevalence of the role model population on the probability that the child smokes. Clearly, defining changes of equivalent magnitude in two completely different variables is a challenge. We take a 'let the data speak' approach and allow both variables to increase by half their standard deviation. This corresponds to a 9.8 percentage point increase in the share of parents who socialize their children at least once a week (from 41.6% to 51.4%)[82], and to a 2.6 percentage point increase in the smoking prevalence rate of the role model population (from 26.3% to 28.9%). We find that these changes cause the likelihood that the child smokes to decrease by 3.9 percentage points and to increase by 2.8 percentage points, respectively. The implication is that parental influences are stronger than social influences in the determination of youth smoking participation. Note that we reach this conclusion without taking into account the direct effects of parental smoking behavior on the probability of youth smoking participation (e.g. due to genetics, mimicking, nicotine addiction from passive smoking, or easier access to cigarettes). Our results suggest that this channel of transmission is also important; the child is more likely to have ever smoked when the PCG has ever smoked. The reported marginal effects, therefore, understate the true effect of the parental channel of transmission.

All regressions in Tables 2 and 3 control for the PCG's education, family income, state cigarette tax, state income, and a wealth of demographic variables.[83] The signs of the estimated coefficients on all these variables are in the expected direction and consistent with empirical findings in the health economics literature. The child is more likely to have ever smoked when the PCG is less educated, when family and state income is low, and when cigarette taxes are low. The PCG is more likely to socialize the child at least once a week when s/he is highly educated, when family and state income is higher, and when cigarette taxes are higher (e.g. because, when taxes are high, smoking by the child entails a higher financial cost for the entire family). From this point on, we do not show estimated coefficients on the control variables, but rather focus on the variables of interest.

[82]Because the parental socialization variable is binary, we calculate its standard deviation (0.19) using its estimated value from the reduced-form regression. To induce an increase in this variable that corresponds to half of its standard deviation, we had to randomly shift PCGs from category 0 to category 1 so that the mean of the variable increases by 9.8 percentage points. This is mathematically equivalent to increasing the probability of socialization for those parents who do not socialize their kids at least once a week from 0 to 0.17 (=(0.19/2)/(1-0.416)).

[83]To save space, we do not present coefficients on the age fixed effects of the PCG and the age, sex, state, race, and religion fixed-effects of the child. However, full results for all models are available on request.

3.6.2 Robustness analysis

Although the baseline specification is already very conservative, there are reasons that induce us to test its robustness to new instruments and controls. First, the performance of our instruments may be impaired by their limited variation; e.g., our measure of PCG exposure to health information does not vary across PCGs who live in the same state and were born in the same year. We could benefit from an exposure measure that is parent-specific and adds individual-level variation to our instrument. Second, our definition of the exposure measure (i.e. as the accumulated sum of anti-smoking articles) imposes the restrictive assumptions that information does not decay and that it has constant returns to scale. More flexible specifications of the exposure measure may be more appropriate. Third, in our baseline specification we do not account for differences in the personality of children or in parenting styles. Both of these have been shown in the psychology literature to vary with children's smoking behavior and the effectiveness of parental socialization efforts (e.g. Huver et al. 2007). Fourth, in our baseline specification we do not control for differences across PCGs in the cost of the parental socialization efforts. We next try to address these issues using a new set of variables which Table 3.6 presents along with some basic descriptive statistics.

Our first exercise aims at introducing individual-level variation in our set of instruments. To do this, we draw from the PSID-CDS-TA a variable that measures how often the PCG reads the newspaper during the week. Assuming that this variable is highly correlated with magazine readership, we interact it with the PCG's information exposure measure to generate a new variable that varies across PCGs. Column 1 of Table 3.7 tests the robustness of the baseline specification to this inclusion. The results are highly robust, and the new instrument appears with a positive and significant coefficient, suggesting that anti-smoking information is more effective at increasing parental socialization efforts when parents have the habit of reading the newspaper often. In comparison to the baseline results, the diagnostic test results are slightly improved (e.g. the probability value of the the ALN test statistic increases to 0.578).

As a further robustness test, we define information exposure to be the average number (instead of the accumulated sum) of anti-smoking articles that the PCG or the child potentially read since age 10. This specification essentially decreases the contributing value of each anti-smoking article by a factor proportional to the age of the PCG, so that (older) PCGs who see a given number of articles over a longer life-span end up with a lower information exposure score than (younger) PCGs who see the same number of articles over a shorter life-span. In other words, the new exposure measure allows the value of information to decay over time. Columns 2 and 3 of Table 3.7 show how the baseline estimates change when we replace the original exposure measure with this new measure, and when we interact it with the frequency the PCG reads the newspaper. Although the results are qualitatively robust to this change and the instrument coefficients are positive and statistically significant, the tests of instrument identification reject the hypothesis that the instruments are valid.

Finally, in Table 3.8 we check the robustness of the baseline estimates to controls that capture personality traits of the child and parenting styles of the PCG (column 1), to controls that capture the cost of PCG socialization efforts (column 2), and to both set of controls together (column 3).

The first set of controls includes: (i) an indicator of whether the child has a tendency to "break rules", which we create by combining a selection of variables documenting "problematic" past behavior (see note of Table 3.6 for the exact definition); (ii) an indicator of high and strict parental control, which we create by combining responses of the PCG to questions about the number of rules s/he impose on the child, whether these rules are strictly enforced, and whether s/he discusses these rules with the child; (iii) an indicator of violent parenting, which flags whether the child is spanked more than 3 times per week; and (iv) an indicator of complete lack of communication among the PCG and the child. Controlling for these variables is potentially important because it may further address the problem of reverse causality between socialization efforts of the PCG and the smoking behavior of the child. Rebellious children or children subject to authoritative parenting may react against parental anti-smoking pressure and may be more likely to smoke when the socialization efforts of the PCG are more frequent. We find that the these controls significantly predict the dependent variables in both the structural and reduced-form equation, while leaving the remaining coefficients almost unaffected. As expected, children who are "rule-breakers" and children who are spanked regularly are more likely to ever smoke and more likely to be socialized by PCGs. In contrast, children who are under strict parental control are less likely to ever smoke and are less often socialized by parents. Finally, children who never discuss any subject with their parents are more likely to ever smoke.

Our measures of the parental socialization cost include: (i) the number of individuals younger than 18 in the family unit; (ii) an indicator of whether the PCG is employed; (iii) the weekly hours that the PCG spends at work; (iv) an indicator of whether the PCG works a regular daytime schedule; and (v) an indicator of whether it takes the PCG over an hour to get to work each way. Once more, we find that, when included in the baseline regression, these variables significantly predict the dependent variables and only result in small quantitative changes in the estimated coefficients of the other variables. Children who live in households with many other children are less likely to smoke and more likely to be socialized by PCGs. Children of working parents are less likely to smoke but also less likely to be socialized, whereas children whose parents work more and regular hours are more likely to smoke and more likely to be socialized.

When we insert in the baseline specification all the new controls together we obtain similar results. All three specifications of Table 3.8 pass the diagnostic tests. In fact, the probability values of the ALM test for specifications (1) and (3) are higher in comparison to that of the baseline specification (0.860 and 0.648, respectively), suggesting that the addition of the controls aids identification.

3.6.3 The mechanisms underlying the transmission process

In the previous sections we found a positive coefficient on the smoking prevalence of the role-model population in both the reduced-form and the structural-form equations. We interpreted these findings as evidence for cultural substitution and conformity. In this section, we scrutinize these results by carrying out two exercises. First, we interact parental smoking status with the prevalence rate of the role-model population to inform our discussion of cultural substitution. Second, we use alternative definitions of the role-model population to inform our discussion on cultural conformity. Both exercises allow us to confirm the links between substitution and conformity and between complementarity and distinction. We conclude this section by further exploring differences in the socialization process between smoker and non-smoker parents.

3.6.3.1 Substitution versus complementarity

Assuming that both smoker and non-smoker parents wish to discourage smoking, cultural substitution entails that all parents should increase anti-smoking socialization efforts in response to the smoking rate in society. The first-stage estimates in columns 1 and 2 of Table 3.9 show that this prediction is supported by the data. We find positive and significant coefficients on all interactions of smoking prevalence and parental smoking status. We read this result to suggest that our extension of the Bisin and Verdier model is more appropriate to use when studying the transmission of traits like smoking than the Bisin and Verdier model in its original version. If smoker parents wished to transmit the smoking culture to their children, like the Bisin and Verdier model would assume, then we would expect to find evidence of cultural complementarity for smoker parents and cultural substitution for non-smoker parents (i.e. positive coefficients on the interaction terms in Table 3.9 for non-smoker parents, and negative coefficients for smoker parents). However, our results suggest otherwise.

As our model predicts, we find similar behavior across smokers and non-smoker parents because there is an exogenous force (anti-smoking information) that has changed people's perceptions of smoking from a 'good' trait to a 'bad' trait. As a result, parents wish to socialize children to be non-smokers, even if they themselves smoke. We follow this logic to develop another implication: we should find evidence of culture complementarity among smoker PCGs exposed to little (or very low) anti-smoking information because, absent other information, these would still consider smoking to be a good trait. In column 3 of Table 3.9 we attempt to test this hypothesis by interacting the parental smoking status with smoking prevalence in the role model population at different quantiles of parental information exposure (quantiles 0-10, 10-50, 50-90, and 90-100).

We note that, in our data there is no PCG who is subject to no anti-smoking information. All PCGs are exposed to some non-negligible level of information. In fact, the PCG at the 10th percentile of the distribution of our exposure variable saw 8 articles, relative to a mean of 12. Therefore, it comes as no surprise that we do not find evidence of cultural complementarity in our results (none of the estimated coefficients carries a negative sign). Nonetheless, there is enough variation in the data to reveal two important patterns that corroborate our theoretical set-up. First, when the smoking prevalence of the role model population increases, all parents respond by increasing their socialization efforts, and their response is larger the more exposed they are to information about the health risks of smoking. Second, this response is not statistically different from zero for smoker parents who are exposed to very low levels of information. These findings encourage us to speculate that, had we observed in our data parents with no or negligible information exposure, we would be able to document a switch in the direction of the relationship between parental socialization and the proportion of smokers in society from positive to negative.

3.6.3.2 Conformity versus distinction

To this point, all the results we have reported are based on the assumption that children derive their role-models from the population that is 20-29 years older. Next, we test whether the probability that the child ever smokes is associated with the smoking prevalence rate of the population of individuals 0-9 and 10-19 years older than the child, and whether that association differs by gender.[84] We present our results in Table 3.10. Estimates in columns 1-4 are based on the total sample, while estimates in columns 5 and 6 are based on data on male and female children, respectively.

The structural form estimates suggest that the probability that the child ever smokes increases with the smoking prevalence rates of all population sub-groups, except for males who are 0-9 years older. The implication is that, in relation to the bulk of the population, children form their smoking decisions based on their needs to achieve assimilation, inclusiveness, and cultural conformity. In relation to the sub-group of young boys, however, the childrens' motive is the exact opposite. In this case, it is their need to generate a sense of distinctiveness from individuals that are part of that group that motivates their smoking decisions. To put it bluntly, our results suggest that young boys operate as anti-role-models. Although this result might seem surprising, one should note that health economic studies on peer effects have not reached a consensus on gender differences in social influence. For example, Nakajima (2007) study peer effects of smoking among school-mates in the US and finds that these are positive and significant within the same gender but statistically negligible across genders. In contrast, Clark and Loheac (2007) study peer effects on different types of risky behavior among friends and school-mates in the US and find significant cross-gender interactions for alcohol use, with young males being more influential than young girls. While we are the first to provide evidence on cultural distinction using smoking data, our evidence complements those presented in the study of ethnic identity formation by Bisin et al.(2013). These authors find that, in neighborhoods in which the share of a given ethnic group is high, the association between the share of the ethnic group and individual ethnic identities is negative.

[84] Naturally, the alternative measures of smoking prevalence rates are correlated with each other, but there is still independent variation in the smoking prevalence rates across the different groups. See Table 3.11 for correlation coefficients and Table 3.6 for means and standard deviations.

3.6.3.3 The link between substitution (complementarity) and conformity (distinction)

In both Tables 9 and 10, the smoking prevalence rates significantly predict the PCGs socialization efforts in the reduced form and the probability that the child smokes in the structural form. Importantly, these effects always run in the same direction, a result which reconfirms the inter-connection between cultural substitution and conformity and between cultural complementarity and distinction, and corroborates Prediction 2 of the model. This result suggests that, because all parents wish to discourage smoking, they will increase their anti-smoking socialization efforts when smoking becomes more popular among all societal groups to which their children will conform, and they decrease their anti-smoking socialization efforts when smoking becomes more popular among young boys from which their children will want to distinguish themselves.

It is worth noting that both our theory and findings contradict those produced by Patacchini and Zenou (2011), although they apply a comparable exercise to identify the cultural transmission mechanisms of preferences for education. Like our assumption that both smoker and non-smoker parents wish to transmit preferences against smoking to their children, Patacchini and Zenou assume that both educated and uneducated parents wish to transmit preferences in favor of education to their children. However, unlike our prediction that cultural complementarity is tied with cultural distinction, these authors predict that complementarity is tied with conformity. As we briefly discuss in Section 3.3, this prediction relies on their assumption that a high prevalence of education in society creates positive externatilities in the effectiveness of parental socialization efforts by decreasing socialization costs (e.g. because more educated neighbors can help a less educated parent to better socialize the children). Their results support their theory. They find that all parents socialize their kids in favor of education, that their socialization effort increases with the prevalence of educational attainment in the population residing in the same neighborhood and, at the same time, the neighborhood education level increases the probability that children acquire education.

3.6.3.4 Other mechanisms

Also relevant is that the theoretical predictions of Patacchini and Zenou rely on the assumption that educated parents are more effective in socializing their children than uneducated parents because they face lower cost of socialization. Our model necessitates no particular assumption about the mechanisms underlying the vertical transmission process of smoker versus non-smoker parents. On the contrary, it allows many mechanisms to be at work at the same time. For example, smoker and non-smoker parents may or may not differ in the effectiveness of their socialization efforts, in their tolerance of having children who smoke, in their perception of the health-risks of smoking, and other dimensions. Whether or not each of these scenaria is true is an empirical question.

The evidence we present in Table 3.9 already shed some light on this issue. We find that never-smoker parents respond to the popularity of smoking in society by increasing their socialization efforts both more than ever-smokers (column 1) and more than current and ex-smokers (column 2). Further, we find that this difference persists at all levels of information exposure (column 3). We present more evidence in Table 3.12. There we show that never-smoker parents also have a higher responsiveness to health information both relative to ever smokers (column 1) and relative to current and ex-smokers (column 2). Both these findings suggest that smoker and non-smoker parents evaluate differently the health risks that their children face when they smoke, and are consistent with existing empirical evidence that non-smokers tend to overestimate the impact of smoking on health (e.g. Viscusi and Hakes 2008). In Table 3.12 we also show that never-smoker parents lower their socialization efforts when they work and increase their socialization efforts when their child is a rule-breaker less than ever-smokers (column 3 and 5, respectively). This may be because, unlike parents who smoke, never-smoker parents are more effective at socializing their kids to be non-smokers by setting the right example and can, therefore, afford to lower their socialization efforts when socialization cost increases or when they have reactive children. Somewhat at odds with this interpretation is our finding that the responsiveness to the socialization cost does not statistically differ between never-smokers and current smokers (column 4), whereas the responsiveness to having a reactive child does not statistically differ between never-smokers and ex-smokers (column 6).

3.6.4 Forward projections of the share of children who ever smoke

As a final exercise, we simulate the dynamics of youth ever-smoking rates over time as described by equation (2) in Section 3.3.5. To do this we use the estimated structural parameters from column 3 of Table 3.8 (i.e. our most restrictive specification), and the mean probability that a child ever smokes that we observe in our sample as the initial condition. This exercise serves two purposes.

First and foremost, it shows that the smoking trait persists in the steady state of the population dynamics, it provides an estimate of the smoking rates at the steady state, and it provides an estimate of the time that smoking rates would need in order to adjust to that steady-state. Second, this exercise provides a framework that we can use to test how different policy regimes can affect the speed of adjustment of the youth smoking rates to equilibrium. We draw the projections in Figure 3.6.

The solid line shows how the youth ever-smoking rates would fare in future generations all else equal. It indicates that the share of youth ever-smokers converges to a steady-state value just above 11 percent.[85] This evidence confirms the long term persistence of the smoking trait, and is consistent with Predictions 3 and 4. Full convergence to that steady-state occurs within five generations (which are on average 25 years apart); i.e. for individuals that are born between 2108-2121. However, 90 percent of the adjustment occurs within only three generations; i.e. for individuals who are born between 2058-2071.

This trajectory can be altered by a social planner via several policy instruments. Here we give the examples of 'reasonable' increases in cigarette taxes and in parental socialization efforts during the lifetime of the children observed in our data. The dotted line shows how the smoking rates would fare under half a standard deviation increase in cigarette taxes (i.e. under a 0.27 dollar increase); and the dashed line shows how they would fare under half a standard deviation increase in the share of parents who socialize their children at least once a week (i.e. an increase of 9.8 percentage points). The former is an example of a policy that targets children directly, while the latter is an example of a policy that targets children indirectly, via parental behavior.

We find that the proposed increase in taxes accelerates the rate of adjustment of youth ever-smoking rates so that 90 percent of the adjustment is achieved within two generations. In comparison, the proposed increase in socialization accelerates the rate of adjustment so that 90 percent of the adjustment is achieved within only one generation. Similarly, the increase in taxes reduces the rate of young smokers at the steady-state by 0.8 percentage points, whereas the increase in socialization reduces it by 1 percentage point. By no means, do we present these results to suggest

[85]Because we use our most restrictive specification to carry out the projections, our results are based on a sample size of 2074. The probability that the child ever smokes in this sample equals 19.6% (as opposed to the 18% reported in Table 3.1). Our results suggest that, in order to reach their equilibrium value, youth smoking rates need to drop by 8.6 percentage points (from 19.6% to 11%).

that subsidizing parental socialization is a preferred tobacco control policy relative to a tax increase. Such a claim would require evidence on the cost of each policy regime. Rather, this evidence demonstrates that a policy maker has the tools to significantly change both the steady-state smoking rates of young people and the speed of adjustment to that steady state. For example, *Healthy People 2020* reports that 19.5 percent of adolescents in grades 9 through 12 smoked cigarettes in 2009 and sets the objective to reduce that percentage to 16 percent by 2020. The results we present in Figure 3.6 suggest that, by using the right policy regime, the government can achieve and even surpass that goal.

3.7 Conclusion

Building on the literature on cultural transmission, we develop a model of smoking dynamics that focuses on the role of parents and social norms; we use novel data to test its theoretical predictions; and we find empirical support. Our chapter advances the literature in several ways.

On the theory front, we extend the seminal work by Bisin and Verdier (2001) to provide a rationale for why traits first emerge, why parents may change the traits they prefer to transmit to their children, and how long-term cultural heterogeneity can be achieved when that occurs. We argue that a cultural trait may emerge when a profit-maximizing industry promotes it, and the way people perceive that trait can be influenced by the flow of related information. Thus, relative to the existing theory, we contribute a framework of analysis to study the transmission mechanisms of a wider variety of cultural traits. Specifically, our model can be used to study (i) cultural traits that already exist and the way people perceive them does not change over time (e.g. preferences on education; trust; religion); (ii) cultural traits that already exist but the way people perceive them changes over time (e.g. preferences on smoking, polluting the environment, or women's rights); and (iii) cultural traits that are brand new (e.g. the culture of social networking). To date, the cultural transmission theory has focused on the traits in the first category.

In developing our model, we introduce new mechanisms to characterize the cultural transmission process. Specifically, we relax a standard assumption in the cultural transmission theory that, when children adopt their traits from society, this happens via a (random or non-random) matching process. This matching process entails that the probability of acquiring a trait always increases with the prevalence of that trait in society. In our model, we allow the industry to affect the direction of this relationship. Borrowing terminology from the literature on identity formation, we formally define the positive relationship between the probability of adopting a trait and the prevalence of that trait in society as cultural conformity. We show that conformity in the social channel of transmission is always tied to substitution in the parental channel of transmission, which posits that parents increase their socialization efforts the more prevalent their preferred cultural trait is in society, and is a cornerstone assumption in the literature. Conformity and substitution will co-exist because parents will lower their socialization efforts when their preferred trait becomes more popular among societal groups to which they know that their children will conform. Correspondingly, we formalize the assumption of cultural distinction to predict the opposite of cultural conformity; i.e. a negative relationship between the probability of adopting a trait and the prevalence of that trait in society. We show that distinction in the social channel of transmission is always tied to complementarity in the parental channel of transmission, which posits that parents decrease their socialization efforts the more prevalent their preferred cultural trait is in society. Distinction and complementarity will co-exist because parents will increase their socialization efforts when their preferred trait becomes more popular among societal groups from which they know that their children will distinguish themselves. We take all our theoretical predictions to U.S. data and we find supporting evidence.

To test our model, we carefully account for the endogeneity of the parental socialization efforts using a novel measure of parental exposure to anti-smoking information as an instrument. We also avoid selection, simultaneity, and exogenous correlation bias in the estimated social effects by measuring smoking prevalence of the role-model population at the state-level, and including a full set of fixed effects in our regression models. Thus, our empirical analysis contributes to the health economics literature causal estimates of parental and social influences on youth smoking participation. Whether smoking behavior is transmitted through parents, role-models, or peers is relevant for designing tobacco control policy and anti-smoking campaigns. If children primarily pick up smoking from their parents, policies that target parental behavior may be more effective at preventing smoking onset than policies that target young people directly. If children primarily pick up smoking from the society, then this implies externalities that can lead to large differences in smoking behavior through social-multiplier effects. On the one hand, social pressure can cause consumption to be sticky in the face of policy instruments; on the other hand, social influences can complement government interventions to prevent smoking initiation among young people. Our results suggest that parental influences are more important than social influences in the transmission of the smoking trait, and they showcase the spread of anti-smoking information as a key instrument to lower smoking rates among young people.

Finally, we demonstrate how the cultural transmission theory can provide an analytical framework which policy makers can use to evaluate the long-term effects of tobacco control policies. Specifically, we use our empirical results to project what will be the steady state rate of youth smoking in future generations. We show that the rate converges to a steady state in which smoking persists. But we also show that a policy maker can affect both the level of smoking at that steady state and the speed of adjustment to that steady state. Our projections suggest that, under the right policy regime, it is possible to achieve the youth smoking rate objectives set by *Healthy People 2020*.

3.8 Figures

Figure 3.1: Smoking prevalence of role-model population of 14 year old children by sex, state, and year

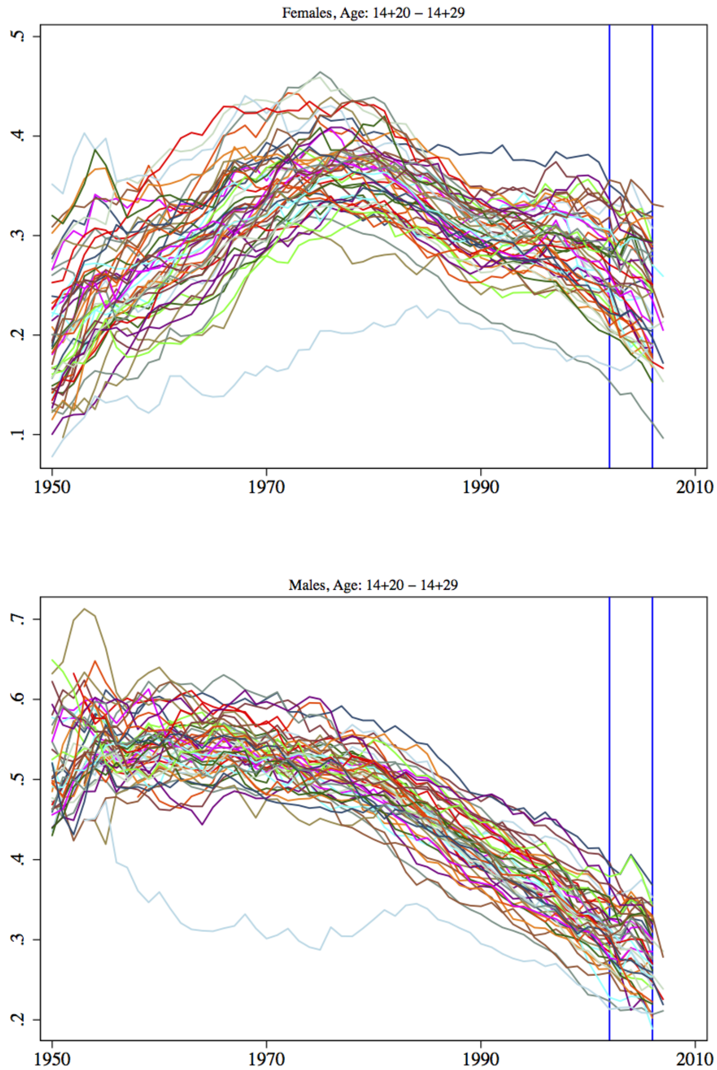


Figure 3.2: Smoking prevalence of role-model population by sex, age, and state

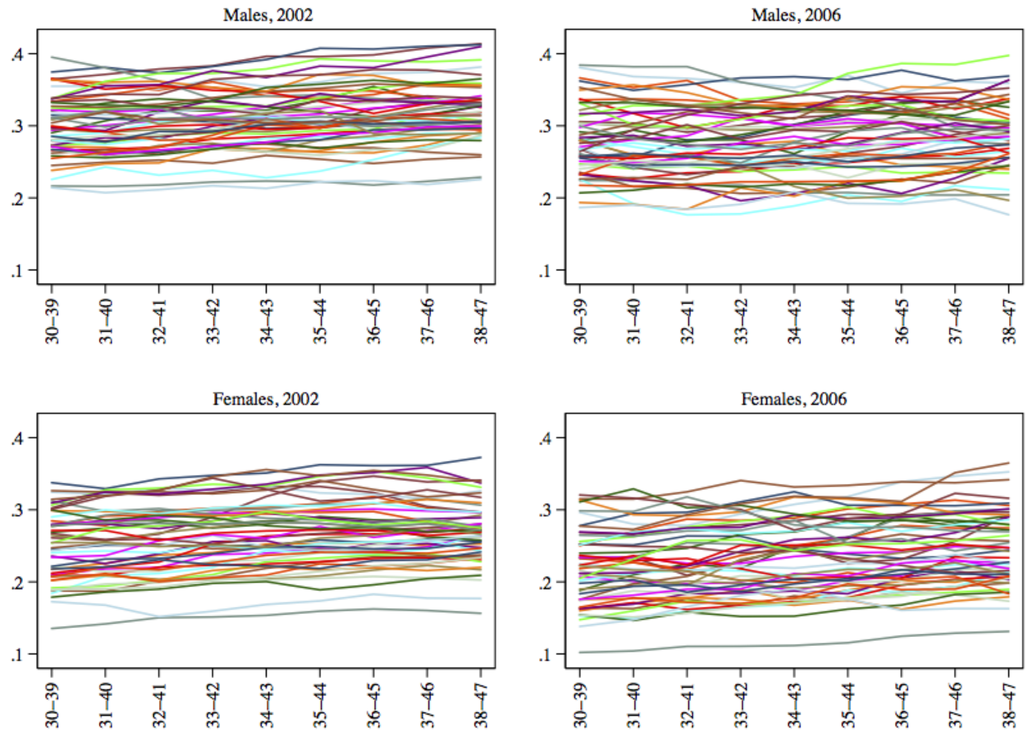


Figure 3.3: Number of published magazine anti-smoking articles by state and year (weighted by state subscription rate to each magazine)

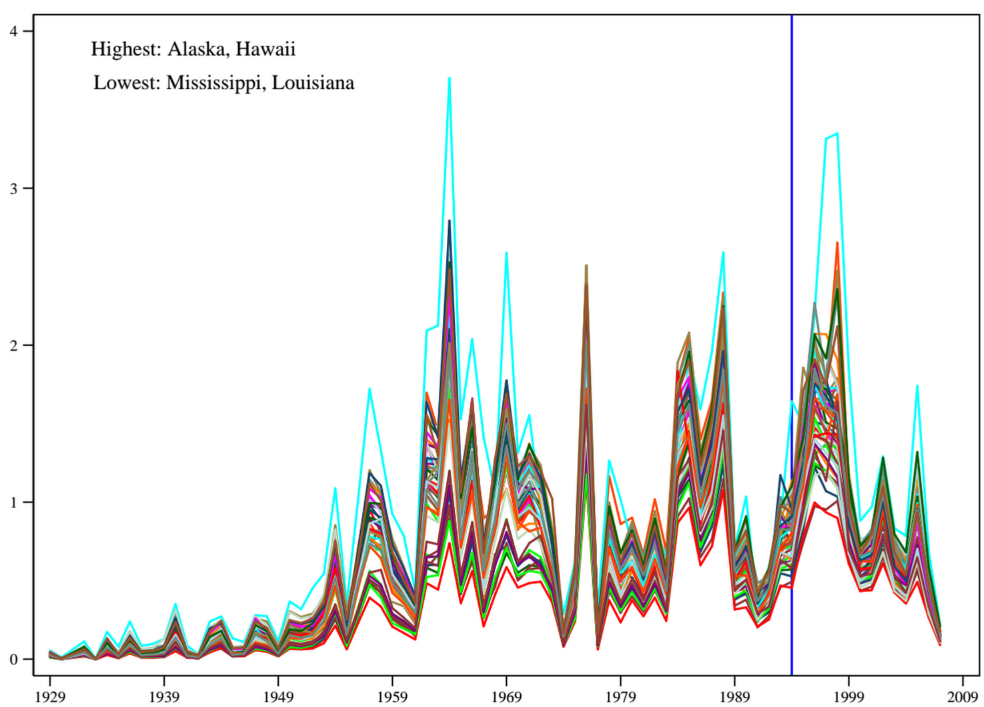


Figure 3.4: Cigarette taxes, by state and year ($)

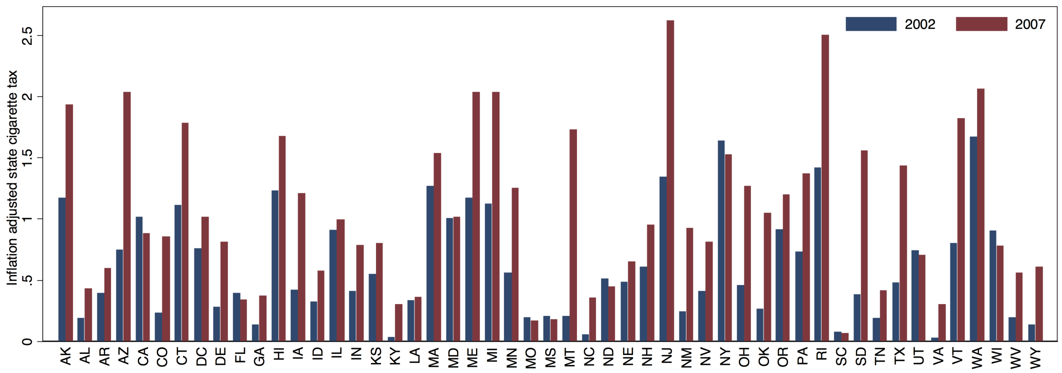


Figure 3.5: Real income per capita, by state and year ($)

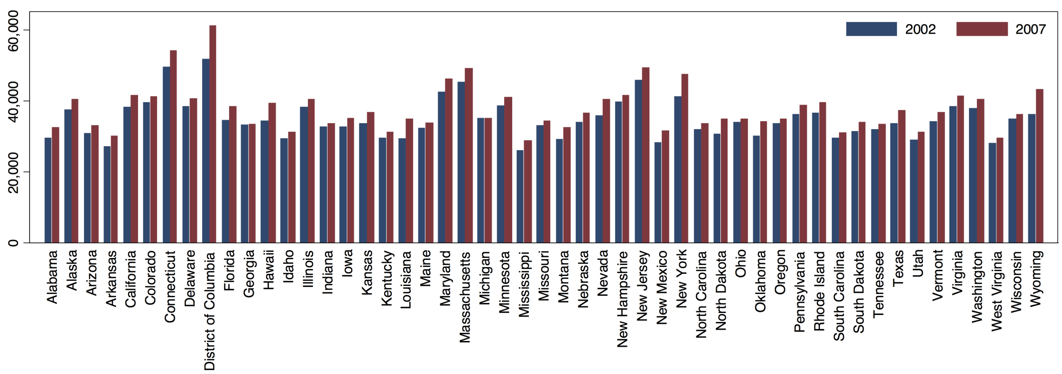
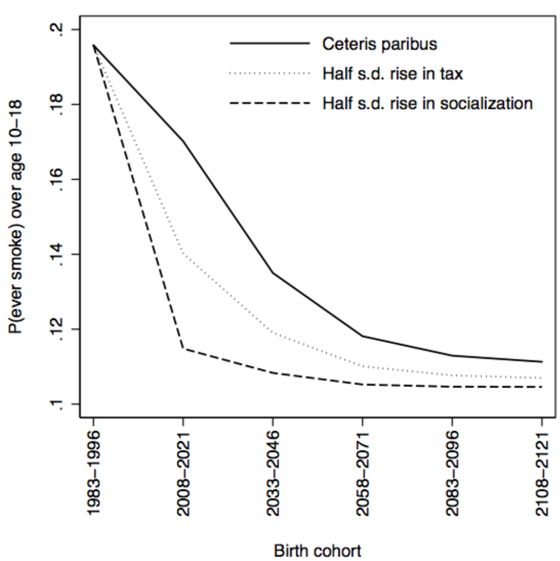


Figure 3.6: Forward projections of youth ever-smoking rates



3.9 Tables

Table 3.1: Weighted means and frequencies of selected variables

Variable	$\bar{x}$/%	[s.d]
Child		
Female	0.50	
Age	13.83	[2.25]
Race		
White, non Hispanic	0.66	
Black, non Hispanic	0.16	
Other	0.18	
Ever smoked regularly	0.18	[0.39]
Acc. sum of anti-smoking articles potentially read since age 10	4.07	[0.05]
Primary Care Giver (PCG)		
Age	41.99	[6.10]
Education	13.02	[2.75]
Relationship to child		
Biological mother	0.93	
Biological father	0.06	
Adoptive mother	0.01	
Stepmother	0.00	
Socializes child against substance use		
Never	0.21	
Once or twice a month	0.42	
Once a week	0.14	
Several times a week	0.13	
Every day	0.10	
Ever smoked regularly	0.45	[0.50]
Acc. sum of anti-smoking articles potentially read since age 10	12.2	[2.78]
St.deviation of anti-smoking articles potentially read since age 10	0.42	[0.07]
Family unit		
Real family income/10000	4.31	[5.22]
State		
Real cigarette tax	1.59	[0.54]
Real state income/10000	3.66	[0.50]
Sm. prevalence of total population 20-29 years older than child	0.26	[0.05]

Observations: 2246.

Table 3.2: Probit regression of the probability that the child has ever smoked

	All categories	Dichotomized: 1='At least'			
	(1)	(2)	(3)	(4)	(5)
Child socialized against substance use					
Once or twice per month	0.351***	0.288***			
	(0.026)	(0.024)			
Once a week	0.256***		-0.025		
	(0.033)		(0.020)		
Several times a week	0.187***			-0.050**	
	(0.033)			(0.022)	
Every day	0.192***				-0.031
	(0.037)				(0.031)
Sm. prevalence of role-model population	3.399***	3.117***	3.598***	3.579***	3.522***
	(1.048)	(1.047)	(1.043)	(1.041)	(1.041)
PCG has ever smoked	0.329***	0.333***	0.337***	0.336***	0.338***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Education of PCG	-0.044***	-0.038***	-0.041***	-0.042***	-0.041***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Family income	-5.453***	-5.433***	-5.121***	-5.100***	-5.127***
	(0.388)	(0.387)	(0.385)	(0.385)	(0.385)
State cigarette tax	-0.930***	-0.919***	-0.900***	-0.902***	-0.901***
	(0.074)	(0.074)	(0.074)	(0.074)	(0.074)
State income per capita	-2.404***	-2.450***	-2.372***	-2.374***	-2.384***
	(0.171)	(0.172)	(0.171)	(0.170)	(0.170)
Pseudo R-squared	0.257	0.256	0.252	0.252	0.252

Controls: Dummies for age of PCG, and age, sex, state, race, and religion of child. Observations: 2246. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.3: IV probit estimates of baseline specification

	Prob(child ever smokes)	Prob(child socialized at least 1/week)
Child socialized at least once/week	-1.461***	
	(0.159)	
Sm. prevalence of role-model population	4.929***	1.309***
	(0.856)	(0.197)
PCG has ever smoked	0.251***	-0.003
	(0.026)	(0.005)
Education of PCG	-0.085***	-0.037***
	(0.005)	(0.001)
Family income	-3.504***	0.471***
	(0.468)	(0.048)
State cigarette tax	-0.656***	0.045***
	(0.083)	(0.0110)
State income per capita	-1.684***	0.148***
	(0.205)	(0.022)
Child info exposure since age 10	-0.025*	-0.012**
	(0.014)	(0.005)
PCG info exposure since age 10		0.018***
		(0.006)
Infrequency of PCG info exposure since age 10		-1.160***
		(0.101)
Amemiya-Lee-Newey minimum χ^2 statistic	0.82 [0.365]	
Hausman χ^2 statistic	252.0 [0.000]***	
Wald χ^2 statistic	40.4 [0.000]***	
χ^2 for joint significance of instruments	165.4 [0.000]***	
Change in prob(child ever smokes) after half s.d increase in:		
Socialization	-0.039	
Sm. prevalence of role-model population	0.028	

Controls: Dummies for age of PCG, and age, sex, state, race, and religion of child. Obs: 2246. Standard errors in parenthesis; p-values in brackets. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.4: Robustness of baseline estimates to alternative methods of estimation

	Linear probability model			Bivariate	Probit
	2SLS	IVLIML	IVGMM	probit	+ Ord. probit
	(1)	(2)	(3)	(4)	(5)
Second-stage					
Child socialization	-0.187***	-0.207***	-0.173**	-1.501***	-0.415***
	(0.063)	(0.066)	(0.071)	(0.019)	(0.147)
Sm. prev. of role-model pop.	1.215***	1.246***	1.153***	4.171***	4.982***
	(0.187)	(0.191)	(0.176)	(0.721)	(1.164)
PCG has ever smoked	0.072***	0.072***	0.073***	0.223***	0.338***
	(0.004)	(0.004)	(0.004)	(0.016)	(0.018)
Education of PCG	-0.017***	-0.018***	-0.016***	-0.086***	-0.077***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.014)
Family income	-0.346***	-0.337***	-0.354***	-3.054***	-4.779***
	(0.048)	(0.049)	(0.048)	(0.297)	(0.406)
State cigarette tax	-0.028***	-0.027***	-0.028***	-0.534***	-0.821***
	(0.009)	(0.009)	(0.008)	(0.059)	(0.079)
State income per capita	-0.072***	-0.070***	-0.072***	-1.487***	-2.344***
	(0.019)	(0.020)	(0.015)	(0.134)	(0.171)
Child info exposure since age 10	0.034***	0.034***	0.035***	-0.022	-0.014
	(0.003)	(0.003)	(0.003)	(0.013)	(0.016)
First-stage					
Sm. prev. of role-model pop.	1.305***	1.305***	1.305***	4.438***	2.984***
	(0.200)	(0.200)	(0.204)	(0.657)	(0.487)
PCG has ever smoked	-0.002	-0.002	-0.002	-0.014	0.011
	(0.005)	(0.005)	(0.005)	(0.015)	(0.012)
Education of PCG	-0.036***	-0.036***	-0.036***	-0.115***	-0.090***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
Family income	0.473***	0.473***	0.473***	1.400***	1.145***
	(0.048)	(0.048)	(0.047)	(0.137)	(0.113)
State cigarette tax	0.046***	0.046***	0.046***	0.180***	0.231***
	(0.011)	(0.011)	(0.011)	(0.033)	(0.027)
State income per capita	0.148***	0.148***	0.148***	0.440***	0.257***
	(0.023)	(0.023)	(0.023)	(0.071)	(0.056)
Child info exposure since age 10	-0.011**	-0.011**	-0.011**	-0.067***	-0.011
	(0.005)	(0.005)	(0.005)	(0.014)	(0.010)
PCG info exposure since age 10	0.016***	0.016***	0.016***	0.081***	0.057***
	(0.006)	(0.006)	(0.006)	(0.016)	(0.011)
Infreq. of PCG info exp. since age 10	-1.161***	-1.161***	-1.161***	-3.717***	-2.505***
	(0.105)	(0.105)	(0.114)	(0.275)	(0.210)

Controls: Dummies for age of PCG, and age, sex, state, race, and religion of child. Observations: 2246. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The child socialization variable in (1)-(4) is the Prob(child socialized at least once/week) and in (5) it is the original ordinal variable that takes values 1-5 to measure how often PCGs socialize children. In (5) we estimated the 1st stage by ordered probit, and the 2nd stage by probit using the hat values from the 1st stage as the socialization variable. We did not adjust the errors.

Table 3.5: Robustness of baseline estimates to alternative definitions of the socialization variable

	Dichotomized: 1='At least'		
	Once or twice/month	Several times a week	Every day
Second-stage: Prob(child ever smokes)			
Child socialization	-1.465***	-2.215***	-3.249***
	(0.324)	(0.150)	(0.183)
Sm. prevalence of role-model population	3.397***	3.809***	5.218***
	(0.862)	(0.718)	(0.652)
PCG has ever smoked	0.266***	0.136***	0.195***
	(0.0345)	(0.0343)	(0.031)
Education of PCG	-0.049***	-0.097***	-0.085***
	(0.004)	(0.004)	(0.003)
Family income	-3.625***	-2.369***	-2.415***
	(0.695)	(0.528)	(0.514)
State cigarette tax	-0.618***	-0.334***	-0.302***
	(0.121)	(0.100)	(0.097)
State income per capita	-1.835***	-1.420***	-1.088***
	(0.282)	(0.219)	(0.234)
Child info exposure since age 10	0.005	0.002	-0.057***
	(0.014)	(0.012)	(0.011)
First-stage: Prob(child socialization)			
Sm. prevalence of role-model population	0.345**	0.752***	1.057***
	(0.175)	(0.168)	(0.120)
PCG has ever smoked	0.012**	-0.024***	0.010***
	(0.004)	(0.004)	(0.003)
Education of PCG	-0.012***	-0.034***	-0.020***
	(0.001)	(0.001)	(0.001)
Family income	0.353***	0.329***	0.083***
	(0.042)	(0.041)	(0.029)
State cigarette tax	0.061***	0.090***	0.048***
	(0.010)	(0.009)	(0.007)
State income per capita	0.033*	-0.007	0.040***
	(0.020)	(0.019)	(0.014)
Child info exposure since age 10	0.030***	-0.0004	-0.019***
	(0.005)	(0.004)	(0.003)
PCG info exposure since age 10	-0.025***	0.015***	0.011***
	(0.006)	(0.004)	(0.003)
Infreq. of PCG info exp. since age 10	-0.214**	-0.620***	-0.379***
	(0.089)	(0.087)	(0.064)

All controls and remaining information, as in Table 3.3.

Table 3.6: Weighted means and frequencies of new controls and instruments

	$\bar{x}/\%$	[s.d]	N
Indicators of child personality and parenting styles			
Child is a 'rule-breaker'	0.31	[0.46]	2232
Child subject to many rules, strictly enforced, rarely discussed	0.10	[0.31]	2232
Child spanked more than 3 times per week	0.01	[0.09]	2232
PCG never discusses any subject with child	0.01	[0.08]	2232
Measures of socialization cost			
Number of children in the family unit	2.10	[1.15]	2088
PCG has a job	0.71	[0.45]	2088
Hours per week PCG typically works	27.5	[20.5]	2088
PCG usually works a regular daytime schedule	0.57	[0.49]	2088
Typically takes PCG over an hour to get to work each way	0.05	[0.21]	2088
Alternative measures of information exposure			
Mean anti-smoking articles PCG potentially read per year since age 10	0.88	[0.13]	2236
Days per week the PCG reads the newspaper	1.96	[1.94]	2236
Alternative definitions of role-mode sm. prevalence			
State-specific smoking prevalence of:			
Total population 0-9 years older than child	0.19	[0.07]	2246
Total population 10-19 years older than child	0.26	[0.06]	2246
Females 0-9 years older than child	0.17	[0.07]	2246
Females 10-19 years older than child	0.23	[0.06]	2246
Females 20-29 years older than child	0.24	[0.06]	2246
Males 0-9 years older than child	0.21	[0.07]	2246
Males 10-19 years older than child	0.30	[0.05]	2246
Males 20-29 years older than child	0.29	[0.05]	2246

Rule breaker: the child did something dangerous, damaged public property, got in a fight, drove drunk or high over 10 times the last 6 months or that the child has been arrested or put in jail more than once to date or that the child has a lot of secrets or hides a lot of things from parents.

Table 3.7: Robustness of the baseline specification to new instruments

	(1)	(2)	(3)
Second-stage: Prob(child ever smokes)			
Child socialized at least once/week	-1.465***	-1.436***	-1.576***
	(0.138)	(0.174)	(0.132)
Sm. prevalence of role-model population	4.376***	4.527***	3.952***
	(0.861)	(0.889)	(0.850)
First-stage: Prob(child socialization)			
Sm. prevalence of role-model population	1.231***	1.245***	1.186***
	(0.197)	(0.210)	(0.211)
PCG info exposure since age 10 (acc. sum)	0.017***		
	(0.006)		
* frequency PCG reads the newspaper	0.0005***		
	(0.0000)		
PCG info exposure since age 10 (mean)		0.290***	0.273***
		(0.082)	(0.079)
* frequency PCG reads the newspaper			0.006***
			(0.001)
Infrequency of PCG info exposure since age 10	-1.167***	-1.206***	-1.181***
	(0.099)	(0.104)	(0.103)
Observations	2,236	2,246	2,236
Amemiya-Lee-Newey minimum χ^2 statistic	1.095	17.0***	17.1***
Hausman χ^2 statistic	252.0***	253.9***	295.0***
Wald χ^2 statistic	53.3***	33.8***	57.6***
χ^2 for joint significance of instruments	216.5***	163.6***	194.0***

Controls: for (1) as in baseline specification; (2) and (3) control for child info exposure since age 10 (mean) instead of (acc. sum).

Table 3.8: Robustness of the baseline specification to new controls

	(1)	(2)	(3)
Second-stage: Prob(child ever smokes)			
Child socialized at least once a week	-1.424***	-1.843***	-1.836***
	(0.179)	(0.095)	(0.108)
Sm. prevalence of role-model population	4.248***	3.535***	2.773***
	(0.894)	(0.760)	(0.791)
Child is a 'rule-breaker'	0.417***		0.389***
	(0.024)		(0.025)
Child subject to many rules, strictly enforced, rarely discussed	-0.593***		-0.562***
	(0.031)		(0.035)
Child spanked more than 3 times per week	0.546***		0.707***
	(0.115)		(0.093)
PCG never discusses any subject with child	0.611***		0.395***
	(0.102)		(0.093)
No. of kids below 18 in the family unit		-0.032***	-0.027***
		(0.010)	(0.010)
PCG has a job		-0.303***	-0.374***
		(0.039)	(0.039)
Hours per week PCG typically works		0.002**	0.003***
		(0.001)	(0.001)
PCG works a regular daytime schedule		0.089***	0.086***
		(0.021)	(0.022)
Takes PCG over an hour to get to work each way		-0.062*	-0.004
		(0.036)	(0.037)
First-stage: Prob(child socialized at least once/week)			
Sm. prevalence of role-model population	1.312***	1.006***	0.979***
	(0.197)	(0.208)	(0.208)
PCG info exposure since age 10	0.017***	0.033***	0.032***
	(0.006)	(0.006)	(0.006)
Infrequency of PCG info exposure since age 10	-1.121***	-1.205***	-1.171***
	(0.101)	(0.105)	(0.105)
Child is a 'rule-breaker'	0.054***		0.060***
	(0.005)		(0.005)
Child subject to many rules, strictly enforced, rarely discussed	-0.109***		-0.100***
	(0.007)		(0.008)
Child spanked more than 3 times per week	0.367***		0.368***
	(0.026)		(0.026)
PCG never discusses any subject with child	-0.038		-0.047*
	(0.028)		(0.027)
No. of kids below 18 in the family unit		0.015***	0.014***
		(0.003)	(0.003)
PCG has a job		-0.177***	-0.184***
		(0.011)	(0.011)
Hours per week PCG typically works		0.003***	0.002***
		(0.000)	(0.000)
PCG works a regular daytime schedule		0.045***	0.046***
		(0.007)	(0.007)
Takes PCG over an hour to get to work each way		0.015	0.025**
		(0.011)	(0.011)
Observations	2,232	2,088	2.074
Amemiya-Lee-Newey minimum χ^2 statistic	0.045	1.299	0.209
Hausman χ^2 statistic	242.7***	358.1***	350.1***
Wald χ^2 statistic	31.0***	90.1***	71.9***
χ^2 for joint significance of instruments	157.8***	139.6***	133.4***

Controls: as in baseline specification.

Table 3.9: Robustness of the baseline specification to new controls

	(1)	(2)	(3)
Second-stage: Prob(child ever smokes)			
Child socialized at least once a week	-1.424***	-1.843***	-1.836***
	(0.179)	(0.095)	(0.108)
Sm. prevalence of role-model population	4.248***	3.535***	2.773***
	(0.894)	(0.760)	(0.791)
Child is a 'rule-breaker'	0.417***		0.389***
	(0.024)		(0.025)
Child subject to many rules, strictly enforced, rarely discussed	-0.593***		-0.562***
	(0.031)		(0.035)
Child spanked more than 3 times per week	0.546***		0.707***
	(0.115)		(0.093)
PCG never discusses any subject with child	0.611***		0.395***
	(0.102)		(0.093)
No. of kids below 18 in the family unit		-0.032***	-0.027***
		(0.010)	(0.010)
PCG has a job		-0.303***	-0.374***
		(0.039)	(0.039)
Hours per week PCG typically works		0.002**	0.003***
		(0.001)	(0.001)
PCG works a regular daytime schedule		0.089***	0.086***
		(0.021)	(0.022)
Takes PCG over an hour to get to work each way		-0.062*	-0.004
		(0.036)	(0.037)
First-stage: Prob(child socialized at least once/week)			
Sm. prevalence of role-model population	1.312***	1.006***	0.979***
	(0.197)	(0.208)	(0.208)
PCG info exposure since age 10	0.017***	0.033***	0.032***
	(0.006)	(0.006)	(0.006)
Infrequency of PCG info exposure since age 10	-1.121***	-1.205***	-1.171***
	(0.101)	(0.105)	(0.105)
Child is a 'rule-breaker'	0.054***		0.060***
	(0.005)		(0.005)
Child subject to many rules, strictly enforced, rarely discussed	-0.109***		-0.100***
	(0.007)		(0.008)
Child spanked more than 3 times per week	0.367***		0.368***
	(0.026)		(0.026)
PCG never discusses any subject with child	-0.038		-0.047*
	(0.028)		(0.027)
No. of kids below 18 in the family unit		0.015***	0.014***
		(0.003)	(0.003)
PCG has a job		-0.177***	-0.184***
		(0.011)	(0.011)
Hours per week PCG typically works		0.003***	0.002***
		(0.000)	(0.000)
PCG works a regular daytime schedule		0.045***	0.046***
		(0.007)	(0.007)
Takes PCG over an hour to get to work each way		0.015	0.025**
		(0.011)	(0.011)
Observations	2,232	2,088	2.074
Amemiya-Lee-Newey minimum X^2 statistic	0.045	1.299	0.209
Hausman X^2 statistic	242.7***	358.1***	350.1***
Wald X^2 statistic	31.0***	90.1***	71.9***
X^2 for joint significance of instruments	157.8***	139.6***	133.4***

Controls: as in baseline specification.

Table 3.10: Probit estimation of the probability that the child is socialized at least once/week

	(1)	(2)	(3)
Second-stage: Prob(child ever smokes)			
Child socialized at least once/week	-1.462***	-1.457***	-0.998***
	(0.159)	(0.161)	(0.208)
Prev. of RMP * PCG never smoked	4.455***	4.154***	
	(0.875)	(0.881)	
Prev. of RMP * PCG ever smoked	5.158***		
	(0.867)		
Prev. of RMP * PCG currently smokes		4.938***	
		(0.870)	
Prev. of RMP * PCG used to smoke		3.966***	
		(0.878)	
Prev. of RMP * PCG never smoked * PCG info exposure q0-10			0.461***
			(0.080)
Prev. of RMP * PCG never smoked * PCG info exposure q10-50			0.295***
			(0.062)
Prev. of RMP * PCG never smoked * PCG info exposure q50-90			0.252***
			(0.056)
Prev. of RMP * PCG never smoked * PCG info exposure q90-100			0.318***
			(0.0541)
Prev. of RMP * PCG ever smoked * PCG info exposure q0-10			0.573***
			(0.083)
Prev. of RMP * PCG ever smoked * PCG info exposure q10-50			0.540***
			(0.066)
Prev. of RMP * PCG ever smoked * PCG info exposure q50-90			0.485***
			(0.057)
Prev. of RMP * PCG ever smoked * PCG info exposure q90-100			0.478***
			(0.057)
First-stage: Prob(child socialized at least once/week)			
Prev. of RMP * PCG never smoked	1.367***	1.361***	
	(0.200)	(0.200)	
Prev. of RMP * PCG ever smoked	1.195***		
	(0.206)		
Prev. of RMP * PCG currently smokes		1.197***	
		(0.206)	
Prev. of RMP * PCG used to smoke		1.119***	
		(0.208)	
Prev. of RMP * PCG never smoked * PCG info exposure q0-10			0.053***
			(0.019)
Prev. of RMP * PCG never smoked * PCG info exposure q10-50			0.078***
			(0.017)
Prev. of RMP * PCG never smoked * PCG info exposure q50-90			0.094***
			(0.016)
Prev. of RMP * PCG never smoked * PCG info exposure q90-100			0.124***
			(0.016)
Prev. of RMP * PCG ever smoked * PCG info exposure q0-10			0.030
			(0.020)
Prev. of RMP * PCG ever smoked * PCG info exposure q10-50			0.061***
			(0.017)
Prev. of RMP * PCG ever smoked * PCG info exposure q50-90			0.096***
			(0.016)
Prev. of RMP * PCG ever smoked * PCG info exposure q90-100			0.086***
			(0.016)
χ^2 test of the equality of the coefficients in first-stage	3.538*	11.64***	237.5***

Observations: 2246. Controls: as in baseline specification, excluding sm. prevalence of role model population.

Table 3.11: Baseline specification under alternative definitions of the role-model population

	Total population				Males	Females
	(1)	(2)	(3)	(4)	(5)	(6)
Second-stage: Prob(child ever smokes)						
Child soc. at least once/week	-1.439***	-1.442***	-1.424***	-1.424***	-2.153***	-1.842***
	(0.163)	(0.163)	(0.161)	(0.165)	(0.183)	(0.094)
Sm. prevalence of:						
total pop. 0-9 years older	-0.083					
	(0.594)					
total pop. 10-19 years older	5.434***					
	(0.892)					
total pop. 20-29 years older	5.366***					
	(0.851)					
males 0-9 years older		-1.088**		-5.766***	-3.917***	-4.909***
		(0.513)		(0.666)	(1.010)	(0.912)
males 10-19 years older		5.543***				
		(0.707)				
males 20-29 years older		3.690***				
		(0.746)				
females 0-9 years older			1.328**	5.114***	7.787***	2.268***
			(0.549)	(0.661)	(2.179)	(0.866)
females 10-19 years older			1.375**			
			(0.689)			
females 20-29 years older			4.145***			
			(0.683)			
First-stage: Prob(child socialized at least once/week)						
Sm. prevalence of:						
total pop. 0-9 years older	-0.023					
	(0.159)					
total pop. 10-19 years older	1.272***					
	(0.208)					
total pop. 20-29 years older	0.153					
	(0.178)					
males 0-9 years older		-0.186		-0.552***	-0.824***	-0.003
		(0.137)		(0.160)	(0.239)	(0.237)
males 10-19 years older		0.087				
		(0.138)				
males 20-29 years older		1.247***				
		(0.173)				
females 0-9 years older			0.329**	0.711***	1.286***	0.594**
			(0.142)	(0.164)	(0.266)	(0.249)
females 10-19 years older			0.428***			
			(0.164)			
females 20-29 years older			0.466***			
			(0.166)			
Observations	2,246	2,246	2,246	2,246	1,058	1,045

Controls: as in baseline specification.

Table 3.12: Correlation coefficients of alternative measures of the sm. prevalence of the role-model population

Measure 1:	Measure 2:	Correlation
Smoking prevalence in total population:		
0-9 years older than child	10-19 years older than child	0.4839
0-9 years older than child	20-29 years older than child	0.6679
10-19 years older than child	20-29 years older than child	0.8355
Smoking prevalence in male population:		
0-9 years older than child	10-19 years older than child	0.3405
0-9 years older than child	20-29 years older than child	0.5592
10-19 years older than child	20-29 years older than child	0.7127
Smoking prevalence in female population:		
0-9 years older than child	10-19 years older than child	0.5732
0-9 years older than child	20-29 years older than child	0.7106
10-19 years older than child	20-29 years older than child	0.8505
Smoking prevalence in population 0-9 years older than child:		
Males	Females	0.9170

Table 3.13: Probit estimation of the probability that the child is socialized at least once/week (first-stage)

	PCG info exposure		PCG has a job		Child is a rule-breaker	
	(1)	(2)	(3)	(4)	(5)	(6)
Interacted with:						
PCG never smoked	0.020***	0.020***	-0.066***	-0.065***	0.046***	0.046***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
PCG ever smoked	0.012**		-0.044***		0.068***	
	(0.006)		(0.008)		(0.007)	
PCG currently smokes		0.013**		-0.056***		0.087***
		(0.006)		(0.009)		(0.009)
PCG used to smoke		0.012*		-0.032***		0.035***
		(0.006)		(0.010)		(0.011)
Observations	2,246	2,246	2,101	2,101	2,238	2,238
χ^2 test	17.7***	22.0***	3.9**	11.0***	4.7**	22.5***

Controls: as in baseline specification, excluding interacted variable.

3.10 Proofs

Proposition 3.1: *Let $v(.)$ be a differentiable, increasing and concave function, and $c(.)$ be a differentiable, increasing and convex function. When interior, a parent's optimal investment choice λ_i^* solves*

$$\frac{v'(\lambda_i^*)}{c'(\lambda_i^*)} = \frac{1}{-u_i(H)\,S(q)}. \tag{14}$$

Proof of Proposition 3.1: The parent invests an amount $\lambda_i^* \equiv \arg\max_{\lambda_0}(1-v(\lambda_i))\,S(q)\,u_i(H) - c(\lambda_i)$ into anti-smoking culturalization. The assumptions on $v(.)$ and $c(.)$ ensure the maximization problem is concave, and so the solution solves the First Order Condition $-v'(\lambda_i)\,S(q)\,u_i(H) - c'(\lambda_i) = 0$, which can be rewritten as (9).⊠

Proposition 3.2: *Performing comparing statics on the optimal investment derived in Proposition 3.1, we obtain the following predictions:*

- *Prediction 1: Parents who perceive* larger *health costs from smoking invest* more *in anti-smoking culturalization.*
- *Prediction 2: We have cultural substitution iff cultural conformity holds, and cultural complementarity iff cultural distinction holds.*

Proof of Proposition 3.2: By total differentiation of (9), we get

$$\frac{d\lambda_i^*}{dH} = \frac{-u_i'(H)}{(u_i(H))^2}\frac{1}{S(q)}\frac{[c'(\lambda_i^*)]^2}{c''(\lambda_i^*)\,v'(\lambda_i^*) - v''(\lambda_i^*)\,c'(\lambda_i^*)} > 0$$

(Prediction 1). We then always have $\frac{d\lambda_{-i}^*}{dq} = \frac{S'(q)}{[S(q)]^2}\frac{1}{-u_i(H)}\frac{[c'(\lambda_i^*)]^2}{c''(\lambda_i^*)v'(\lambda_i^*) - v''(\lambda_i^*)c'(\lambda_i^*)} > 0$ iff $S'(q) > 0$, and the effect of an increase in smoking prevalence on the likelihood a child becomes a smoker can be expressed as $(1-v(\lambda_i^*))\,S'(q)$ (Prediction 2). ⊠

Proposition 3.3: *In a subgame perfect equilibrium, we have $S(0, \theta^*(0)) > 0$, which ensures a heterogeneous distribution of traits in steady state. Moreover, both cultural substitution and cultural conformity may obtain.*

Proof of Proposition 3.3: In a subgame perfect equilibrium, the firm's objective function can be written as:

$$\max_{\theta} Q(q, \theta) - \kappa(\theta),$$

where $Q(q, \theta) \equiv [q(1 - a(q, \theta)) + (1 - q)(1 - d(q, \theta))] S(q, \theta)$ denotes the proportion of children who become smokers given a proportion q of adult smokers. whose solution $\theta^*(q)$ solves the FOC:[86]

$$[q(1 - a(q, \theta)) + (1 - q)(1 - d(q, \theta))] \frac{\partial S(q, \theta)}{\partial \theta} = \kappa'(\theta)$$

In order to understand when cultural distinction vs. conformity arise, we need to study the sign of: $\left.\frac{dS}{dq}\right|_{\theta^*} = \frac{\partial S}{\partial q} + \left.\frac{\partial S}{\partial \theta}\right|_{\theta^*} \frac{\partial \theta^*}{\partial q}$, where $\frac{\partial \theta^*}{\partial q} = \frac{[d(q) - a(q) - qa'(q) - (1-q)d'(q)] \frac{\partial S(q,\theta)}{\partial \theta} + [q(1-a(q)) + (1-q)(1-d(q))] \frac{\partial^2 S(q,\theta)}{\partial \theta \partial q}}{c''(\theta) - [q(1-a(q)) + (1-q)(1-d(q))] \frac{\partial^2 S(q,\theta)}{\partial \theta^2}}$. While the denominator of the expression for $\frac{\partial \theta^*}{\partial q}$ is always positive, the numerator can admit both signs. Hence cultural substitution obtains if and only if the numerator takes a large enough negative value. ⊠

Proposition 3.4: *In a steady state, there always exists a fraction of non-smokers. Moreover, under cultural complementarity the smoking habit always persists, while under cultural substitution it persists as long as $S(0) > 0$.*

Proof of Proposition 3.4: When $q = 0$, we have $\dot{q} = (1 - d(0)) S(0)$. Under cultural substitution we have $d(0) = 0$ and so $\dot{q} > 0$ iff $S(0) > 0$. Under cultural complementarity we have $d(0) > 0$ and so $\dot{q} > 0$ iff $S(0) > 0$ and $d(0) < 1$. When $q = 1$, we have $\dot{q} = (1 - a(1)) S(1) - 1$. Under cultural substitution we have $a(1) > 0$ and so $\dot{q} < 0$ always. Under cultural complementarity we have $a(1) = 0$ and so $\dot{q} < 0$ iff $S(1) < 1$. By Proposition 3.2, we have $S' < 0$ under cultural complementarity, and so $S(0) > 0$ and $S(1) < 1$. ⊠

[86] Our assumptions ensure that the second order condition satisfies

$$SOC : \left[q\left(1 - a^1(q)\right) + (1 - q)\left(1 - d^0(q)\right)\right] \frac{\partial^2 S(q, \theta)}{\partial \theta^2} - \kappa''(\theta) < 0$$

and so we have a concave problem whose solution can be recovered through the FOC.

References

Algan, Yann, and Pierre Cahuc. 2010. "Inherited Trust and Growth." *American Economic Review*, 100(5): 2060–2092.

Aquilino, W. S. 1994. "Interview mode effects in surveys of drug and alcohol use: a field experiment." *Public Opinion Quarterly*, 58: 210–240.

Avenevoli S. and Ries Merikangas K. 2003. "Familial influences on adolescent smoking." *Addiction,* 98 (Suppl. 1): 1–20.

Becker G. S. and Murphy K. M. 1988. "A Theory of Rational Addiction." *Journal of Political Economy*, 96 (4): 675–700.

Bisin, Alberto and Thierry Verdier, 2000. "Beyond the Melting Pot: Cultural Transmission, Marriage and the Evolution of Ethnic and Religious Traits." *Quarterly Journal of Economics, 115: 955–988.*

Bisin, Alberto and Thierry Verdier, 2001. "The Economics of Cultural Transmission and the Dynamics of Preferences." *Journal of Economic Theory*, 97: 298–319.

Bisin, Alberto and Thierry Verdier, 2011. "The Economics of Cultural Transmission and Socialization." in *Handbook of Social Economics, Jess Benhabib, Alberto Bisin, and Matt Jackson (eds), Elsevier Science.*

Bisin, Alberto, Topa, G., Verdier,T., 2004. "Religious intermarriage and socialization in the United States." *Journal of Political Economy*, 112: 615–664.

Bisin Alberto, Patacchini Eleonora, Verdier Thierry and Zenou Yves. 2013. "Bend It Like Beckham: Ethnic Identity and Integration," Unpublished manuscript

Borjas, George J. 1992. "Ethnic Capital and Intergenerational Mobility." *Quarterly Journal of Economics*, 107(1): 123–150.

Chern, Wen S, Loehman, Edna T, and Yen, Steven T, 1995. "Information, Health Risk Beliefs, and the Demand for Fats and Oils." *Review of Economics and Statistics*, 77(3): 555–564.

Christopoulou, R., Lillard, D. R. 2013. "Is Smoking Behavior Culturally Determined? Evidence from British Immigrants." *NBER Working Paper* No. 19036.

Clark, A.E., and Etile, F., 2006. "Don't give up on me baby: spousal correlation in smoking behavior." *Journal of Health Economics* 25 (5), 958–978.

Clark, Andrew E. and Youenn L. 2007. "It wasn't me, it was them! Social Influence in risky behavior by adolescents." *Journal of Health Economics.* 26: 763–784.

Dohmen Thomas, Armin Falk, David Huffman, Uwe Sunde, 2012. "The Intergenerational Transmission of Risk and Trust Attitudes." *Review of Economic Studies*, 79(2): 645–677.

Fernandez, Raquel, and Alessandra Fogli. 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics*, 1(1): 146–177.

Fletcher Jason, M. 2010. "Social Interactions and Smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects", *Health Economics*, 19: 466–484.

Freedman David A. and Jasjeet S. Sekhon. 2010. "Endogeneity in Probit Response Models." *Political Analysis*, 18(2): 138–150.

Gaviria, Alejandro and Raphael, Steven. 2001. "School-Based Peer Effects And Juvenile Behavior," *Review of Economics and Statistics*, 83(2): 257–268.

Göhlmann, Silja, Christoph M. Schmidt, and Harald Tauchmann. 2010. "Smoking Initiation in Germany: The Role of Intergenerational Transmission." *Health Economics*, 19(2): 227–242.

Harris, Jeffrey E. and Gonzalez Lopez-Valcarcel B. 2008. "Assymetric peer effects in the analysis of cigarette smoking among young people in the United States, 1992-1999" *Journal of Health Economics*. 27: 249–264.

Heckman James J. 1978. "Dummy Endogenous Variables in a Simultaneous Equations System." *Econometrica*, 46(6): 931–959.

Huver Rose M. E., Engels Rutger C. M. E, Van Breukelen Gerard, De Vreis Hein. 2007. "Parenting style and adolescent smoking cognitions and behavior." *Psychology and Health*, 22(5): 575–593.

Jellal Mohamed and Wolff Francois-Charles. 2002. "Cultural evolutionary altruism: theory and evidence", *European Journal of Political Economy*, 18: 241–262.

Lillard, Dean. 2011. "Keeping it in the Family? If Parents Smoke Do Children Follow?" *Journal of Applied Social Science Studies*. 131: 277–286.

Lillard Dean R. and Sfekas Andrew. 2013. "Just passing through: the effect of the Master Settlement Agreement on estimated cigarette tax price pass-through." *Applied Economics Letters*, 20(4): 353–357.

Loureiro, Maria L., Anna Sanz-de-Galdeano, and Daniela Vuri. 2010. "Smoking Habits: Like Father, Like Son, Like Mother, Like Daughter?" *Oxford Bulletin of Economics and Statistics*, 72(6): 717–743.

Lundborg, P., 2006. "Having the wrong friends? Peer effects in adolescent substance use." *Journal of Health Economics* 25 (2), 214–233.

Manski, C.F., 1993. "Identification of endogenous social effects: the reflection problem." *Review of Economic Studies* 60 (3), 531–542.

Manski, C.F., 2000. “Economic analysis of social interactions.” *Journal of Economic Perspectives* 14 (3), 115–136.

Nakajima, Ryo. 2007. “Measuring Peer Effects on Youth Smoking Behaviour.” *Review of Economic Studies*, 74(3): 897-935.

Namoro Soiliou and Rania Roushdy. 2008. “Intergenerational Transmission of Fertility Preferences: A Test of the Cultural Substitution Assumption,” Working Paper 352, University of Pittsburgh, Department of Economics.

Newey W. K. 1987. “Efficient estimation of limited dependent variable models with endogenous explanatory variables.” *Journal of Econometrics*, 36: 231–250.

Nichols Austin. 2011. “Causal Inference for Binary Regression with Observational Data.” CHI11 Stata Conference 6, Stata Users Group.

Orphanides, A., Zervos, D., 1998. “Myopia and addictive behavior.” *Economic Journal* 108 (446), 75–91.

Orphanides, A., Zervos, D., 1995. “Rational addiction with learning and regret.” *The Journal of Political Economy* 103 (4), 739–758.

Patacchini Eleonora and Zenou Yves. 2011. “Neighborhood Effects and Parental Involvement in the Intergenerational Transmission of Education.” *Journal of Regional Science*, 51(5): 987–1013.

Pollak, R. A. 1970. “Habit formation and dynamic demand functions.” *Journal of Political Economy*, 78: 745–763.

Pollak, R. A. 1976a. “Habit formation and long-run utility functions.” *Journal of Economic Theory*, 13: 272–297.

Pollak, R.A., 1976b. “Interdependent preferences.” *American Economic Review* 66 (3), 309–320.

Poutvaara P. and Siemers L-H.R. 2008. “Smoking and Social Interaction”. *Journal of Health Economics*. 27: 1503–1515.

Powell L. M. and Chaloupka F.J. 2005. “Parents, public policy, and youth smoking.” *Journal of Policy Analysis and Management*, Volume 24, Issue 1, p. 93–112.

Powell Lisa M., John A. Tauras, Hana Ross, 2005. “The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior.” *Journal of Health Economics*, 24 (5): 950–968.

Robson, Arthur and Larry Samuelson. 2011. “The Evolutionary Foundations of Preferences.” in *Handbook of Social Economics, Jess Benhabib, Alberto Bisin, and Matt Jackson (eds), Elsevier Science.*

Royal College of Physicians of London (RCPL), 1962. Smoking and Health: Summary and Report of the Royal College of Physicians of London on Smoking in Relation to Cancer of the Lung and other Diseases, London: Pitman Medical Publishing Co. Ltd.

Sáez-Martí María & Sjögren Anna, 2008. "Peers and Culture," *Scandinavian Journal of Economics*, 110(1): 73–92.

Stigler G. J. and Becker G. S. 1977. "De Gustibus Non Est Disputandum." *American Economic Review*, 67 (2):76–90.

U.S. Department of Health, Education and Welfare (USDHEW). 1964. "Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service", *Public Health Service Publication* No. 1103.

Viscusi, W. Kip & Hakes, Jahn K. 2008. "Risk Beliefs And Smoking Behavior," *Economic Inquiry*, 46(1): 45–59.

Viscusi, W. Kip and Hersch Joni. 2011. "Tobacco Regulation Through Litigation: The Master Settlement Agreement." in Regulation versus Litigation: Perspectives from Economics and Law. Daniel P Kessler (editor), University of Chicago Press, 71–101.

Appendix

A1 Additional Figures and Tables

Table A1

Campaign contribution limits

To any candidate committee (per election[1])	To any national party committee (per year)	To any PAC, state/local party, or other political committee (per year)	Aggregate total
Pre-BCRA			
$1,000	$20,000	$5,000	$25,000 per year
2010 Cycle[4]			
$2,400, subject to aggregate limit	$30,400 per party committee, subject to aggregate limit	$10,000 to each state or local party committee (Levin funds) $5,000 to each PAC or other political committee, subject to aggregate limit	• $115,500 per two-year election cycle as follows: • $45,600 per cycle to candidates; and • $69,900 per cycle to all national party committees and PACs (of which no more than $45,600 per cycle can go to PACs)
2008 Cycle[4]			
$2,300, subject to aggregate limit	$28,500 per party committee, subject to aggregate limit	$10,000 to each state or local party committee (Levin funds) $5,000 to each PAC or other political committee, subject to aggregate limit	• $108,200 per two-year election cycle as follows: • $42,700 per cycle to candidates; and • $65,500 per cycle to all national party committees and PACs (of which no more than $40,000 per cycle can go to PACs)
2006 Cycle[4]			
$2,100, subject to aggregate limit	$26,700 per party committee, subject to aggregate limit	$10,000 to each state or local party committee (Levin funds) $5,000 to each PAC or other political committee, subject to aggregate limit	• $101,400 per two-year election cycle as follows: • $40,000 per cycle to candidates; and • $61,400 per cycle to all national party committees and PACs (of which no more than $40,000 per cycle can go to PACs)
2004 Cycle			
$2,000, subject to aggregate limit[2]	$25,000 per party committee, subject to aggregate limit	$10,000 to each state or local party committee (Levin funds)[3] $5,000 to each PAC or other political committee, subject to aggregate limit	• $95,000 per two-year election cycle as follows: • $37,500 per cycle to candidates; and • $57,500 per cycle to all national party committees and PACs (of which no more than $37,500 per cycle can go to PACs)

Reproduced from: http://www.opensecrets.org/bigpicture/limits.php

[1] Primary and general elections count as separate elections.

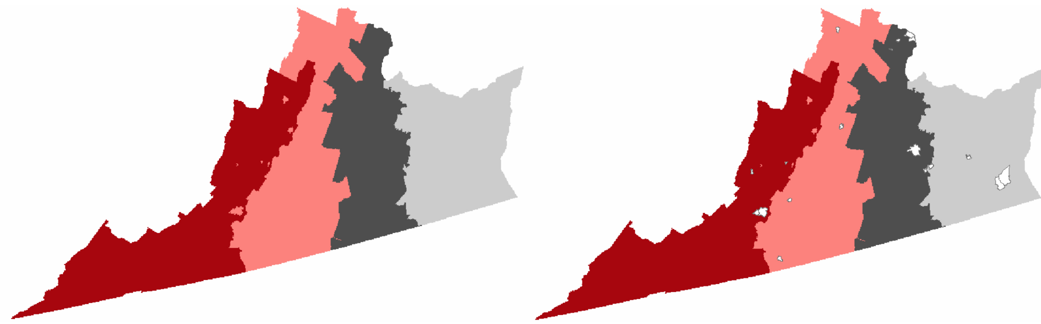
[2] BCRA's individual contribution limits are higher to candidates facing wealthy opponents financing their own campaigns.

[3] Levin funds also can come from corporations and labor unions if allowed by state law.

[4] Under BCRA, some individual contribution limits were indexed for inflation and adjust accordingly every cycle.

[5] Multicandidate committees are those with more than 50 contributors, that have been registered for at least six months and (with the exception of state party committees) have made contributions to five or more federal candidates.

Figure A2: Land elevation in Virginia



Notes: Counties are divided into four quartiles on land elevation (Q1: light grey, Q2: dark grey, Q3: light red, Q4: dark red). In the second map, 100% urban counties are marked in a white box on the map.

Figure A3: Plot of the residuals from a reduced-form regression of turnout on the number of ISPs

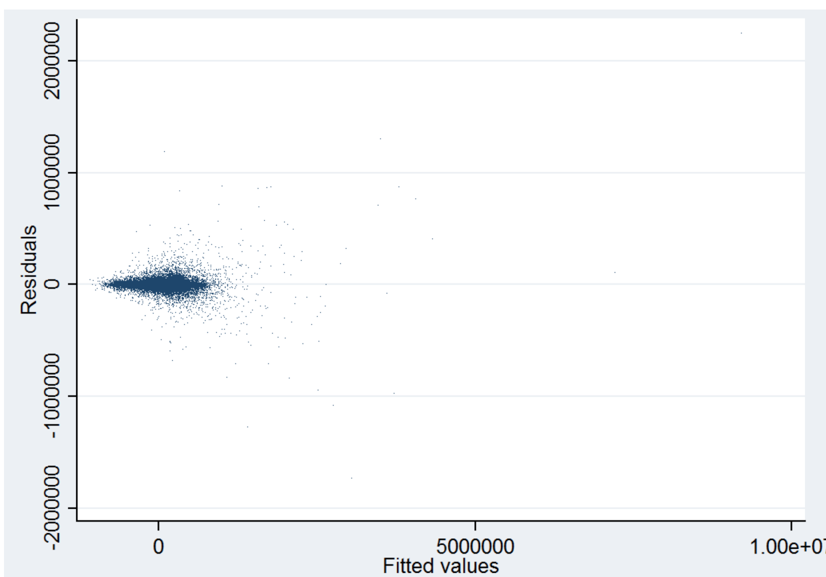


Table A4
Robustness analysis of the effect of broadband on survey data measures

	Know. index	Pro-abortion	Pro-gay	Liberal index	Online donations	
ISPs	0.474***	0.491**	0.374**	0.457**	0.417	
	(0.161)	(0.229)	(0.15)	(0.215)	(0.308)	
Obs.	5,103	11,349	20,201	8,546	2,178	
R^2	0.323	0.172	0.246	0.276	0.098	
	TV	**Internet**	**Cable TV**	**Non-cable TV**	**CBS**	**FOX**
ISPs	-0.606***	0.440**	-0.657**	-0.0913	-0.0754	-0.146
	(0.220)	(0.215)	(0.283)	(0.274)	(0.135)	(0.147)
Obs.	18,969	18,969	9,529	9,529	9,529	9,529
R^2	0.081	0.166	0.039	0.074	0.047	0.027

Notes: The analysis is identical to the ones repored in Tables 12-14, but now adding state-time dummies to the list of controls, as well as county controls. * indicates significance at the 10 percent level, ** indicates significance at the 5 percent level, *** indicates significance at the 1 percent level.

A2 Cable vs ADSL broadband Internet

The history of the Internet can be traced back to Arpnet, a 1960s computer network of the U.S. Department of Defense. In the early 1970s, the private sector – in particular AT&T – was given the opportunity to take over this new technology (Sunstein 2007). However, it is only two decades later that the corporate world received these calls for privatization with enthusiasm. In 1992, the National Science Foundation started supervising the commercialization of the Internet. From then on, government funding was slowly withdrawn, until the Internet was fully privatized in 1995 (Greenstein 2004).

In the early days of the commercial Internet age, residential customers could only connect through low-speed dial-up connections. It would however be wrong to assume this was due to firms' inability to provide high-speed service. In fact, the three technologies which would later dominate the broadband market had already been fully developed by the early 1990s (T1-lines, ADSL, CM).[87] As it turns out, a complex web of market, technological, regulatory and geographical limitations

[87] The word 'Asymmetric' refers to the fact that downstream (i.e. browsing) speed is higher than upstream (i.e. sending) speeds (Grubesic and Murray 2004).

have shaped the broadband map as we know it today.

One reason telephone companies were slow to deploy DSL was their fear of cannibalizing their lucrative commercial T1-lines business (e.g. Greenstein 2004). The spread of DSL was also hampered by the fact that it could only operate in areas where the distance between key components of the infrastructure is small enough (household, central office, network backbone).[88] In 1999, DSL could only potentially reach 44% of the U.S. population (Faulhaber 2002). For example, many rural and high-income suburban areas, as well as a large part of the Southern United States, could not readily get broadband access through DSL.

Federal regulation – in particular the Telecommunications Act of 1996 (TA96) – also made it unattractive for incumbent telephone companies (e.g. Verizon, BellSouth) to enter the broadband market.[89] First, TA96 mandated that high-cost customers be served at the same prices as low-cost customers (Hazlett and Bittlingmayer 2003). Second, TA96 required incumbent firms to lease DSL facilities to their competitors on an unbundled basis and at a predetermined rate (FCC, 2005, pp.19-20).[90] The direct effect of this regulation was to discourage investment in telecommunications infrastructure (Kahn et al. 1999; Hazlett and Havenner 2003). It also created an environment where non-incumbent telephone firms could cheaply provide service on incumbents' lines (Kahn et al. 1999; Hazlett and Havenner 2003). In practice, it also made it difficult for customers to assign blame when broadband service was unsatisfactory (Rosenbush 2001).[91]

In contrast with ADSL, cable companies' technology for broadband required a hefty investment

[88] ADSL only involved a relatively minor hardware cost, through the installation of Digital Subscriber Line Access Multiplexers (DSLAMs) in central offices (Ferguson 2001).

[89] In particular, see Section 706 and the Computer III regime of TA96. A number of subsidy programs were also instigated by the government, again primarily at the federal level (Goolsbee 2002, Wallsten 2005). The only one that has been documented to be somewhat successful is the Universal Service Fund's Rural Health program, which was targeted towards networking projects in schools / libraries in rural / low income areas (Flamm 2005, 2011).

[90] Rates were set by each individual state through rights-of-way laws (ROWs). These state-level laws determine rates for building broadband infrastructure on both publicly- and privately-owned land. According to one estimate, the cost of abiding with ROWs can add up to 20% of the total cost of providing broadband servic. The impact of these rates can be particularly acute in rural areas, where there often are more poles per mile than households. Securing rights to this infrastructure is often a difficult and time-consuming process that discourages private investment. For instance, the application process is more streamlined in some states (e.g. Connecticut, New York). (e.g. zoning rules, regulation of utility poles). See http://www.broadband.gov/plan/6-infrastructure/#_edn3

[91] In 2003, legislators became concerned with the low broadband penetration of telephone companies. This led to an overhaul of the regulatory framework set by TA96 through a series of reforms. In February 2003, the FCC eliminated the DSL line sharing rules which forced incumbents to only sell the high-speed proportion of their lines. From there on, competitors would thus have to pay for the entire local loop or strike a commercial agreement with the carrier to share a loop. In August 2005, the FCC moved further in this direction by putting broadband DSL services on regulatory parity with cable services (Hazlett and Caliskan 2008). These changes were positively received by ILECs, which aggressively deployed DSL starting in 2003. By 2008, the change was visible with DSL and CM controlling roughly equal shares of the broadband residential market (Robinson and Weisman 2008).

in upgrading infrastructure (see Section 1.3.1). This initially put them at a competitive disadvantage relative to telephone companies.[92] However, TA96 had little *direct* impact on the operations of cable operators, whose activities it classified as an "information service." This indirectly encouraged cable companies to enter the broadband market, by giving them a regulatory advantage over incumbent telephone companies. As an added advantage, broadband networks would help cable companies compete better with satellite-based television systems that were threatening their core television operations (Fabrikant and Schisel 2001).[93] Cable firms responded to these incentives by upgrading their one-way TV networks for two-way data transmission, spending over $10 billion per year starting in 1996 (i.e. $200 per home passed).

A3 Pew survey questions

The following questions are used to construct variables on media consumption, liberal values, and campaign donations:

Media: *How do/did you get most of your news about the presidential election campaigns? From television, from newspapers, from radio or from magazines or from the Internet?*

TV follow-up: *Did you get most of your news about the election campaigns from network TV news (ABC, NBC, CBS), from local TV news, or from cable news networks (CNN, Fox News, MSNBC)?*

Pro-choice: *Do you think abortion should be ... legal in all cases, legal in most cases, illegal in most cases, or illegal in all cases?*

Pro-gay: *Do you ... strongly favor, favor, oppose or strongly oppose ... allowing gays and lesbians to marry legally?*

Campaign donations: *Thinking about the past 12 months, have you contributed money to a political candidate or party, or any other political organization or cause?*

Online vs. Offline donating: *Did you make those contributions on the internet... or did you make those contributions offline, say, in person, by phone or through the mail... or have you made contributions both on the internet and offline?*

[92] As for cable companies, they were not required to provide access to their networks to competitors. This position was latter reaffirmed by the Supreme Court in *NCTA v. Brand X Internet Services* (2005).

[93] By using a device known as a statistical multiplexer, an HFC network may carry a variety of services, including analog TV, digital TV, telephony, and high-speed data (Hazlett and Bittlingmayer 2003, Starks 2013).

A respondent can volunteer up to two answers to the Media question, and 5 to 8 answers on the TV follow-up question (depending on the survey). Note that CNBC is mentioned as a possible answer choice in all but one of the surveys that ask the TV follow-up question. To have a consistent measure of cable TV consumption, I base it on the first 5 TV channels reported other than CNBC.[94] For the pro-choice and pro-gay questions, answers are on a 4-point scale, which I collapse them into 2 categories: legal/illegal (pro-choice), and favor/oppose (pro-gay).

I now report questions used for constructing the political knowledge index, arranged by survey. Some questions have open-ended answers, in which case I report in parentheses all valid answer choices. When the survey allows multiple answers, I report in bold the correct answers to the questions. To facilitate the comparison of surveys, I mention next to each question in which other survey (if any) this question also appears. The number of these questions varied between 6 and 17 across surveys.

February 2007 (1,420 Respondents)

Now I would like to ask you about some people who have been in the news recently. Not everyone will have heard of them. If you don't know who someone is, just tell me and I'll move on. Can you tell me who [...] is?

1. Hillary Rodham Clinton (Correct answers: Senator, Bill Clinton's wife, former First Lady, Presidential candidate, Democratic leader)

2. Harry Reid (correct answers: Senate Majority Leader, senator, Democratic leader)

3. Nancy Pelosi (correct answers: Speaker of the House of Representatives, congresswoman, Democratic leader, San Francisco politician) (also in March 2007)

4. Condoleezza Rice (correct answers: Secretary of State, member of Bush's Cabinet, Bush advisor) (also in Feb 2008)

5. Barack Obama (correct answers: Senator, presidential candidate, Democratic leader)

6. Arnold Schwarzenegger (correct answers: Governor of California, actor)

7. Robert Gates (correct answers: Secretary of Defense, member of Bush's Cabinet, Bush advisor, former CIA official) (also in March 2007; Aug 2007; Dec 2008)

[94] Only 2.4% of respondents report watching CNBC to access news about political campaigns.

8. Lewis "Scooter" Libby (correct answers: vice president's former top aide, worked for Dick Cheney, White House aide, accused of divulging name of CIA operative Valerie Plame, on trial)

9. Will you tell me who the Vice-President of the United States is? (correct answers: Dick Cheney, Richard Cheney)

10. Can you tell me the name of the President of Russia? (March 2007; August 2007) (correct answer: Vladimir Putin)

11. Can you tell me the name of the current Governor of your state? (All states excluding DC) (Correct answer depends on state code)

12. Can you tell me the name of the current Mayor of your city? (Only DC) (correct answer: Adrian Fenty)

13. Do you happen to know which political party has a majority in the U.S. House of Representatives? (also in Feb 2008; Dec 2008) (correct answer: Democrats)

14. The Chief Justice of the Supreme Court is John Roberts. Can you tell me if he is generally considered a liberal, a moderate, or a conservative? (Also in Dec 2008) (correct answer: conservative)

15. Has Hillary Rodham Clinton announced that she is running for president in 2008, or has she not announced that she is a candidate? (correct answer: Yes, has announced)

16. And now please tell me, what is the name of the former mayor of New York City who is being mentioned as a possible Republican presidential candidate in 2008? (correct answer: Rudy Giuliani)

17. Since the start of military action in Iraq, about how many U.S. military personnel have been killed? To the best of your knowledge, have there been [...] troop deaths? (Also in Feb 2008; Dec 2008) (correct MCQ answer: Around 3,000)

March 2007 (955 respondents)

1. Do you happen to know which political party has a majority in the U.S. House of Representatives? (also in August 2007; Feb 2008; Dec 2008) (correct MCQ answer: The Democratic party)

2. Can you tell me who is the President of Russia? Is it… (Also in Feb 2007; August 2007) (correct MCQ answer: Vladimir Putin)

3. Can you tell me who is the Vice-President of the United States? Is it... (correct MCQ answer: Dick Cheney)

4. Is Nancy Pelosi. . . (Also in Feb 2007) (correct MCQ answer: The Speaker of the House of Representatives)

5. Is Robert Gates. . . (Also in Feb 2007; Aug 2007; Dec 2008) (correct MCQ answer: The U.S. Secretary of Defense)

6. Followers of the two major branches of Islam are seeking political control in Iraq. One branch is the Shi'a, whose members are known as Shiites. Can you name the other one? Is it . . . (Also in March 2007; Feb 2008) (correct MCQ answer: The Sunni)

August 2007 (940 respondents)

1. Which popular online website sponsored a Democratic presidential debate in July with CNN? Was it... (correct MCQ answer: YouTube)

2. What religion is Republican presidential candidate Mitt Romney? Is he ... (correct MCQ answer: Mormon)

3. Which of the following Democratic presidential candidates is Hispanic? (correct MCQ answer: Bill Richardson)

4. Followers of the two major branches of Islam are seeking political control in Iraq. One branch is the Shi'a, whose members are known as Shiites. Can you name the other one? Is it . . . (Also in March 2007; Feb 2008) (correct MCQ answer: The Sunni)

5. Do you happen to know the name of the current Speaker of the U.S. House of Representatives? Is it. . . (also in Dec 2008) (correct MCQ answer: Nancy Pelosi)

6. Is Robert Gates. . . (Also in Feb 2007; March 2007; Dec 2008) (correct MCQ answer: The U.S. Secretary of Defense)

7. Can you tell me who is the President of Russia? Is it. . . (Also in Feb 2007; March 2007) (correct MCQ answer: Vladimir Putin)

8. Do you happen to know which political party has a majority in the U.S. House of Representatives? (also in March 2007; Feb 2008; Dec 2008) (correct MCQ answer: Democrats)

9. Since the start of military action in Iraq, about how many U.S. military personnel have been killed? To the best of your knowledge, have there been [...] troop deaths? (Also in Feb 2008; Dec

2008) (correct MCQ answer: Around 3,500)

February 2008 (926 respondents)

1. Is Condoleezza Rice... (correct MCQ answer: The U.S. Secretary of State)

2. Followers of the two major branches of Islam are seeking political control in Iraq. One branch is the Shi'a, whose members are known as Shiites. Can you name the other one? Is it ... (Also in March 2007; Aug 2007) (correct MCQ answer: The Sunni)

3. Since the start of military action in Iraq, about how many U.S. military personnel have been killed? To the best of your knowledge, have there been [...] troop deaths? (Also in Aug 2007; Dec 2008) (correct MCQ answer: Around 4,000)

4. Do you happen to know what state U.S. Senator John McCain represents in Congress? Is it... (correct MCQ answer: Arizona)

5. Do you happen to know the name of the current majority leader of the U.S. Senate? Is it... (correct MCQ answer: Harry Reid)

6. Do you happen to know which political party has a majority in the U.S. House of Representatives? (also in March 2007; August 2007; Dec 2008) (correct MCQ answer: Democrats)

7. What is the name of the talk show host who has campaigned for Democratic presidential candidate Barack Obama? Is it... (correct MCQ answer: Oprah Winfrey)

8. Is Hugo Chavez the President of... (correct MCQ answer: Venezuela)

9. Can you name the chairperson of the Democratic National Committee? Is it... (correct MCQ answer: Howard Dean)

10. Do you happen to know who is chairman of the U.S. Federal Reserve Board? Is it... (correct MCQ answer: Ben Bernanke)

11. Which of the following recently declared its independence from Serbia? (correct MCQ answer: Kosovo)

December 2008 (959 respondents)

1. Do you happen to know which political party has a majority in the U.S. House of Representatives? (also in March 2007; August 2007; Feb 2008) (correct MCQ answer: Democrats)

2. To what cabinet position has Barack Obama recently nominated Hillary Clinton? Is it... (correct MCQ answer: Secretary of State)

3. Since the start of military action in Iraq, about how many U.S. military personnel have been killed? To the best of your knowledge, have there been [...] troop deaths? (Also in Aug 2007; Feb 2008) (correct MCQ answer: Around 4,200)

4. Is Nicolas Sarkozy the President of... (correct MCQ answer: France)

5. Is Guantanamo Bay ... (correct MCQ answer: The location of a U.S. naval base)

6. Do you happen to know the name of the current Speaker of the U.S. House of Representatives? Is it... (Also in Dec 2007) (correct MCQ answer: Nancy Pelosi)

7. Is Robert Gates... (Also in Feb 2007; March 2007; Aug 2007) (correct MCQ answer: The U.S. Secretary of Defense)

8. The Chief Justice of the Supreme Court is John Roberts. Can you tell me if he is generally considered a liberal, a moderate, or a conservative? (Also in Feb 2007) (correct MCQ answer: Conservative)