Three Essays in Applied Microeconomics

by

Sarah Elaine Taylor

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2013

Doctoral Committee:

Professor Yusuf Can Masatlioglu, Co-Chair
Professor Mel Stephens, Jr., Co-Chair
Professor Yan Chen
Professor Stephen W. Salant

To my parents, for their continual support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

Growing Up in a Steel Town: Early-Life Pollution Exposure and Later-Life Mortality[1]

1.1 Introduction

Scientists have conducted thousands of studies evaluating the health impacts of air pollution. This concentrated research effort is not surprising; evidence suggests that pollution from a variety of sources—e.g., trucks and automobiles (Finkelstein *et al.*, 2004), energy generation (Lave and Freeburg, 1973), and industrial processes such as steel production (Pope, 1989)—can have important detrimental effects for human morbidity and mortality. Careful studies that document the effects of pollution are crucial for the formulation of appropriate policy responses.

A prevailing theory is that exposure to pollutants is damaging for physical development—with the natural implication that such health deficits could subsequently lead to increased mortality long after an individual is no longer being exposed to pollution. There are well founded concerns that poor early-life conditions might have a "long reach," affecting health in important ways in later-life.[2] For example, in her well-known book, *When Smoke Ran Like Water,* Davis (2002) presents forceful arguments for the proposition that childhood exposure to steel production in towns like Donora had profound negative consequences for the later-life health of individuals.[3]

[1]This research was supported in part by an NIA training grant to the Population Studies Center at the University of Michigan (T32 AG000221). I am also grateful for use of the services and facilities of the Population Studies Center at the University of Michigan, funded by NICHD Center Grant R24 HD041028. I thank Professor Seth Sanders of Duke University for providing me with access to the Duke SSA/Medicare data set, and also for helpful discussions. I would also like to thank Martha Bailey, Dan Black, David Card, Yan Chen, Ryan Kellogg, Yusufcan Masatlioglu, Stephen Salant, Mel Stephens, Evan Taylor, and Lowell Taylor for additional feedback. Any and all errors or misunderstandings in my paper are my own.

[2]The "long reach" idea is linked closely to the work of Barker (1995), and plays an important role in work by Fogel (2004). A recent relevant study on the "Barker hypothesis" is Almond (2006). Among the other papers that study the childhood origins of health are Case, *et al.* (2002), Case, *et al.* (2005), and Hayward and Gorman (2004).

[3]Donora was the location of the infamous "killing fog of 1948." From October 26 to October 31 an air inversion trapped particulate matter from local metal production facilities in Donora, leading to a dramatic spike in mortality (Davis, 2002).

As Davis (2002) indicates, there are good reasons to believe that early-life exposure to pollution may cause long-lasting health deficiencies. For example, childhood exposure to particulate matter (PM) has been shown to lead to deficits in lung functioning (Gauderman *et al.*, 2004). Still, most of the literature on the health impacts of air pollution focuses on the contemporaneous relationship between pollution exposure and health outcomes for individuals (e.g., Chay and Greenstone, 2003b, and Currie and Walker, 2011). There is little work that seeks to establish a link between early-life pollution exposure and health problems that develop over a lifetime or present in later-life. This is a potentially significant gap in the literature. Efforts to explore long reach effects are important for developing a more complete scientific understanding of the impact of pollution. Findings about the lifetime impacts might also be important to policy makers, who need to be able to appropriately evaluate the costs of pollution.

The lack of research exploring the long reach impact is likely due to the difficulties of assembling data that have measures of early-life pollution exposure *and* sufficiently long time periods to examine post-exposure health impacts leading to early mortality.[4] I undertake the first such large-scale study, focusing on early-life exposure to pollution from steel production in small cities in Pennsylvania in the early twentieth century.[5] I find that later-life mortality is impacted negatively by early-life exposure to steel production emissions; evidence on the conditions that may affect the amount of exposure further support the "long reach" hypothesis.

The key to this study lies in the history of the Pennsylvania steel industry. At the beginning of the twentieth century, a large share of U.S. steel production was in Pennsylvania, some of which came from steel production facilities located in towns throughout the Commonwealth. (There were also, of course, thousands of towns that had no steel mills.) Steel production in relatively small cities was still commonplace in the early 1930s. During World War II, there was a vast increase in U.S. steel production with the increase in international demand; this coincided with a substantial concentration of production in large integrated facilities in Pittsburgh and other major steel-producing locations, like Clairton and Donora.[6] Plant closures occurred in smaller and less efficient facilities, such as those in small towns (DiFrancesco, *et al.*, 2010). Indeed, records from 1954 show that most small-town steel plants that operated in the early 1930s were closed down by that year.[7]

[4]While there appears to be no studies linking early-life pollution exposure to older-age mortality, there is research in economics that links early-life conditions to later-life health outcomes. One example is Almond's (2006) examination of *in utero* exposure to the 1918 influenza pandemic for later-life health outcomes. A second example is Maccini and Yang's (2009) study of the relationship between rainfall in the year of birth to later-life outcomes for individuals born in Indonesia.

[5]"Small cities" (also termed "towns" in this paper) are those with below-median birth populations in study cohorts.

[6]By 1940, the Pittsburgh area had 42% of U.S. steel production capacity (Warren, 1973).

[7]Although steel production in Pennsylvania towns was minimal by 1954, there was continued production in a small number of locations (American Iron and Steel Institute, 1954).

My focus is on a large sample of individuals who were born in towns in Pennsylvania from 1916 through 1927. Note that these cohorts were aged 27 to 38 in 1954, at which time most small-town steel production had ceased (American Iron and Steel Institute, 1954). The goal of the study is to see if individuals born in steel towns—who were exposed to pollutants from coal combustion and metallic exhaust as children and possibly into young adulthood—had higher levels of mortality in older age than individuals who were born in non-steel towns. In order to conduct this analysis, it is necessary to have a plausible comparison set; individuals born in steel towns are compared to individuals born in "neighboring" non-steel towns, i.e., to individuals born in non-steel towns in the same county. Because the research design looks at *within county* variation, the analysis focuses exclusively on towns, rather than cities. The reason is that there are very few counties in Pennsylvania that had more than one larger city.

The analysis relies on unique data, the Duke Medicare/SSA Data Set. The starting point of these data is the administrative files of the Medicare Part B program, which include date of birth (used for Medicare eligibility determination) and date of death (used for the purpose of terminating benefits) for people aged 65 and older in the United States. These data do not include place of birth, but they were matched to the Numerical Identification Files of the Social Security Administration, which have "town or county of birth" for most individuals. As described below, the data are reasonably complete for the cohorts I study, 1916–1927. The data include only individuals aged 65 and older, as 65 is the age of eligibility for Part B benefits.

The elements of the dataset are place of birth, zip code of residence in older age or death (among those who are deceased), gender, and race. I merged these records to additional data. First, I used Pennsylvania historical records to determine the town-level location of steel production facilities in the early twentieth century. Second, I matched the older-age residence zip code to the median income in the zip code to form a crude proxy of lifetime prosperity for individuals in my sample. I thereby assembled complete records for more than 780,000 white individuals born in Pennsylvania, 1916–1927. Of this sample, approximately 390,000 were born in "towns"—steel towns and non-steel towns. Below I show that these two comparison groups are highly comparable along observable dimensions.

I find that among individuals in my sample, those who were born (and potentially continued living) in steel towns have significantly higher mortality rates than those born in non-steel towns. The higher mortality is particularly pronounced for individuals who were born in steel towns that had relatively high levels of steel production capacity in the early 1930s. Also, the detrimental long-run impacts of pollution appear to be larger for people born in low-elevation locations, perhaps because pollution exposure was more severe in valleys than in high-elevation towns. Key relationships pertain for born women and men.

Of course, a major concern with my empirical work is the potential for omitted variable bias, as discussed in Chay, Dobkin, and Greenstone (2003). An example is the possibility that localities with metal production tended to be more prosperous than other locations, which could have led to increased labor-force participation, higher wages, and possibly better health care and education. These are factors which would typically improve lifetime well-being and in turn, health outcomes. I discuss this issue when interpreting my findings.

The paper proceeds in three additional sections:

In Section 2 I provide a brief critical review of existing work on the health impacts of TSPs (Total Suspended Particulates), and fine and coarse PM, a particularly dangerous subset of "traditional" TSPs.[8] The primary point that emerges from this review is the absence of papers that evaluate the consequence of early-life exposure to later-life mortality.

In Section 3 I provide a historical overview of the steel industry. I discuss the process by which steel was manufactured and the potentially harmful chemicals that were released during production.

In Section 4 I turn to statistical evidence on the association between early-life exposure to steel pollutants and later-life mortality.

Section 5 concludes.

1.2 Literature

The approaches to studying health impacts of pollution are varied—reflecting differences in expertise and methodology across disciplines. Much of the relevant work falls within the following three overlapping strands of literature:

First, there are many studies in the broad area of public health and environmental sciences that establish associations between exposure to pollutants and morbidity or mortality. A prominent example is the pioneering Harvard Six Cities study (Dockery, *et al.,* 1993), which shows a strong and robust association between city-level air pollution and mortality (caused by cancer and cardiopulmonary disease).

Second, researchers in medicine and epidemiology seek to establish pathophysiological links between pollution and health outcomes. There is experimental work involving animals, and there is

[8] Traditionally TSPs are defined as airborne particles or aerosols that are less than 100 micrometers. In 1987, as a new health standard, the U.S. EPA replaced the TSP with an indicator for only coarse particulate matter, PM_{10}—that is, 10 micrometers or less. Again, in 1997, a stricter standard for particles less than 2.5 micrometers was established, holding fine particulate matter, $PM_{2.5}$, as the dangerous particle of interest (Fierro, 2000).

research exploring physiological impacts through natural means—examining the health of surrounding animal populations or vegetation.[9] Other research focuses on epidemiological evidence regarding pollution and specific hypothesized physiological deficits among humans. Among fascinating work in this field is research demonstrating an association between exposure to fine PM and deficits in lung function development among children (Gauderman, *et al.*, 2004).

Third, economists have made a distinctive contribution by focusing on potential threats to the validity of causal inferences drawn from statistical associations between pollution and health outcomes. Economists often implement novel identification strategies that aim to plausibly establish causality. An example is an analysis by Chay and Greenstone (2003a). These authors motivate their work by arguing that pollution is not randomly assigned to individuals, hence the need to find quasi-random variation for estimating the impact of pollution on health. They proceed by using such variation—generated by the 1981-1982 recession—and find that a reduction in location-specific TSPs generates a substantial decline in the infant mortality rate.

1.2.1 Approaches and Challenges for Studying Health Effects of Particulate Air Pollution

As noted above, a vast literature is concerned with the health consequences of TSP emissions, as examples: those released into the environment as the result of industrial processes, traffic, or other sources. Pope and Dockery (2006) provide an extensive and enlightening review of the extant literature on the health consequences of pollution. In the Pope-Dockery taxonomy, most studies fall into three categories:

First, there are many studies that focus of short-term pollution exposure and mortality. These studies include analyses of severe pollution episodes such as the famous October 1948 "killing fog" in Donora, Pennsylvania (see Davis, 2002, and references therein). While these events are dramatic, systematic analysis is difficult. It is especially difficult to evaluate the extent to which excess mortality is the consequence of harvesting, whereby most mortality is among individuals who would have died soon in any case. For my purposes—the study of early-life pollution exposure and later-life health—such events are unlikely to be useful unless they are truly dramatic.[10]

Second, there are many prominent studies that focus on long-term particulate exposure. These

[9]For instance, Somers, *et al.* (2004) examine the impact of air pollution on heritable mutation rates in birds and rodents, while Wellenius, *et al.* (2003) examine the impact of pollution on myocardial ischemia in dogs.

[10]Thus, it will surely be worthwhile to conduct long-term health studies for individuals who were exposed to radioactive iodine in the aftermath of the 1986 Chernobyl nuclear reactor explosion. (For work examining shorter-term impacts among children, see Almond, Edlund, and Palme, 2009.) Similarly, perhaps exposure to volcanic magmatic gases could be studied for the purposes of determining the long-term impacts of exposure to such elements as aluminum and rubidium (Durand, *et al.,* 2004).

studies include the famous Harvard Six Cities study (Dockery, *et al.,* 1993) and the American Cancer Society (ACS) prospective cohort studies (Pope, *et al.,* 1995). There have been many follow-up papers and re-analyses of the data collected for these studies. This research stream is particularly notable because the follow-up time is impressive (over 16 years in some cases). Still, it is worth noting that the time frame is not nearly long enough to capture "long reach" effects sufficient to investigate any association between exposure in the early stages of life and older-age mortality.

Third, there are studies for which the "time scale" of exposure varies. In some instances such variation might be reasonably thought of as quasi-random. An example is a 13-month shutdown of a steel mill in Utah Valley, which occurred because of a labor dispute (Pope, *et al.,* 1992). Again, though, none of the studies reviewed by Pope and Dockery (2006) evaluate such variation as "long-run effects."

It is finally worth noting that most of the studies mentioned here do not account for the fact that pollution exposure is not randomly assigned. This is a major concern, as emphasized by Chay and Greenstone (2003a). Other researchers, using this careful insightful, have produced meaningful results heeding the advice of Chay and Greenstone, for example, Currie and Walker (2011).[11] Beyond the desire to use quasi-random variation, one must assess how to deal with variables which make a difference for outcomes (such as lifetime well-being) that cannot be directly observed. I discuss such challenges in my empirical section.

1.3 The Steel Industry

To set up the empirical analysis below, it helps to briefly consider the history of iron and steel production, such as early twentieth century integration of production facilities (both in terms of technology and business). Furthermore, an assessment of the technologies used in these plants gives insight into chemical emission.

1.3.1 Technologies and Production Process

The process of making steel, at the most basic level, involves inputs of energy and iron ore. By the early 20th century, the basic process was standard: iron ore was transformed into pig iron, and then refined into steel. Some facilities existed separately for the purpose of melting and reshaping, such as rolling mills, which produce plates or quality flat and long products (Beer, *et al.,* 1998).

In their study of energy-efficient steel making, Beer, *et al.* (1998) produce a thorough survey of

[11] This paper is also a good source for referencing additional efforts along these lines.

the history of the steel and iron industry. In the young market, charcoal was the source fuel. Coke (made from coal) was introduced to the iron market in 1718, and by 1790 it made up 90% of pig iron production (Beer, *et al.,* 1998).

As Beer, *et al.,* describe, several processes had emerged by 1900—perhaps the most famous of which was was the Bessemer Process, patented by Bessemer in 1855. In this process, the metal is melted by the heat of the oxidation of the carbon and the other impurities in the refining process. Cold air is blown through a refractory acid-lined vessel in the Bessemer converter. No additional fuel was required other than the initial 1 ton coal per 1 ton steel consumed. The Bessemer Process was used widely in early twentieth century plants, including those in Pennsylvania.[12] In addition, there was widespread use of the open hearth furnace, which was developed in France (Beer, *et al.,* 1998).

Warren (1973) notes that at the end of the 1800s, the steel industry consisted of a large number of relatively small firms that would buy raw materials on the open market. Pig iron was purchased by production facilities that were only capable of refining from the secondary product. Iron ore and coke, much of which came from Connellsville (Warren, 2001), was procured by pig iron furnaces. In 1903, for efficiency reasons, steel companies began to build by-product coke ovens in the production facilities for integrated production of pig iron and steel, beginning in Pennsylvania in the Pittsburgh area. Warren (1973) suggests that this change in coking was particularly beneficial to Pittsburgh as they could send coking coal down the Monongahela. However, the establishment of linkages between sites in the Commonwealth was slow and it wasn't until 1918 when great headway was made toward integration—when U.S. Steel put up the world's largest by-product coke oven on a site north of Clairton Works, a supply of coal was barged daily down the Monongahela to Pittsburgh (and also gas to Edgar Thomas, Homestead, Duquesne, and Clairton Works).

Pittsburgh and concentrated plants along the confluence of the rivers had a clear advantage in the steel market. However, many production facilities, including those in small towns, survived until the early 1930s. Warren (1973) suggests that with changes in the industry, the market tended to weed out poorly located or inefficient mills and furnaces. By 1954, there was a substantial decline in production at non-central, non-integrated mills located in the small towns of Pennsylvania.

[12]Note here that the"Bessemer Process" could also include the adapted Thomas process or the basic Bessemer Process, which used only basic refractory lining in the Bessemer converter.

1.3.2 Exposure to Pollution from Steel Production

The dangers from exposure to steel and iron production have been widely studied. Coal burning in iron and steel production leads to the release of CO_2, and the process of iron smelting releases CO (Climate Leaders, 2003). Among other symptoms, exposure to CO_2 can lead to asthma and increased rates of cancer (Jacobson, 2008). Exposure to CO can lead to wide-ranging clinical symptoms from cardiovascular to respiratory and neurobehavioral effects at even low concentrations; unconsciousness and death can result after prolonged or acute exposure to high concentrations (Fierro, *et al.*, 2001).

Fierro (2000) notes that additional metallic and gaseous emissions are produced through smelting and steel mills, emitting the most common size of particulate matter, PM_{10} and $PM_{2.5}$. Coarse particulate matter is released in the form of solids as a product of coal burning; fine particulate matter, gaseous particles, are derived from coal combustion and additional chemical reactions from metal processing in these plants. These fine particulates can embed themselves deep in the upper respiratory tract and possibly present harm to organs. PM_{10} has been associated with increased hospital admissions for COPD, asthma, and lower respiratory tract infections, including bronchitis and pneumonia, in elderly patients (Fierro, 2000). Fierro (2000) also points to the strong association between $PM_{2.5}$ and cardiac disease.

Researchers are obtaining a more thorough understanding of how specific compounds in emissions from these integrated steel mills and other coal-burning facilities, such as separate blast iron furnaces and steel mills, affect health through compiled medical studies. A recent report on the toxicological properties of coal emissions showed that coal-burning power plants produce hazardous air pollutants which cause irritation and tissue damage to the eyes, skin, and breathing passages at high levels of exposure (Billing, 2011). Billing additionally suggests that exposure to these pollutants can cause *latent diseases* that can develop over many years and may be a contributing factor to such fatal conditions as heart disease and brain impairments. In a 1989 report on steel mills prepared by the Radian Corporation for the EPA, it is noted that substances of concern to public health not only include coke oven emissions from coal-burning, but also heavy metal emissions (e.g., copper, cadmium, and chromium). Chromium, for example, has been shown to cause damage to nasal passages and, in long run studies (lending support to the long reach hypothesis), has been linked to lung cancer (Radian Corporation, 1989).

One of the most important studies that provides direct evidence about this issue is the Utah Valley "natural experiment" (Pope, 1989). The study shows that during the closure of the steel mill, children's hospital admissions were substantially lower than when the mill was operating. This was particularly true for bronchitis and asthma. Similar findings pertained for adults, though the relationship was not as strong. Pope (1989) also gives links to other literature on pollution from steel. I do not discuss that literature further here, other than to note, again, the literature does not focus on the long-reaching effects of exposure to coal-burning pollution.

1.4 Early-Life Exposure to Pollution from Steel Production and Older-Age Mortality

I address the "long reach" hypothesis—that exposure to pollution at young ages has a negative impact on mortality in later-life. As discussed above, research such as Gauderman, *et al.* (2004), suggests that health deficits to lung functioning can occur at young ages—potentially leading to higher morbidity and mortality at older ages. Pope and Dockery (2006) note that that air pollution exposure likely causes increased cardiopulmonary morbidity, which could appear as higher mortality in older age. This backdrop motivates the question proposed for empirical investigation: "Do individuals who likely had higher levels of exposure to steel production emissions early in life have relatively higher levels of old-age mortality?" My data allow for investigation of this question for mortality post age 65.

1.4.1 Data.

The key to providing evidence about this research question is unique data, the "Duke SSA/Medicare" data, which match complete Medicare Part B records with Social Security records via the Numerical Identification Files (NUMIDENT) of the Social Security Administration. Black *et al.* (2012) indicate that for cohorts born in 1916 and after, these data cover approximately 85% of the population.[13] Location of birth is supplied at the county or town level; Black *et al.* (2012) find that approximately 80% of records are at the town level. I use only data that include location of birth at the town level. The data also include location by zip code at age 65 or date of death (for those who are deceased), and they include gender and race. The records extend through 2002, so it is possible to analyze rates of mortality for people aged 65-75 for cohorts born 1916 through 1927. The relevant variable for "survival" is the rate of living to 75 given the individual has lived to 65.[14]

I merge the Duke SSA/Medicare data to historical records that provide locations of pollution sources. Data on steel production in Pennsylvania comes from two primary sources. The first source

[13] Black, *et al.,* (2012) provide additional details of data construction.

[14] In addition, one could in principle study the incidence of disability for anyone enrolled prior to age 65, though the data have not yet been used for that purpose.

is from the Secretary of Internal Affairs of the Commonwealth of Pennsylvania (1903).[15] The second source is from an independent historical source published by the American Iron and Steel Institute (1930). This later source is especially valuable for my purposes. It gives the location of all steel production plants in Pennsylvania as of 1930, a year in which individuals in my study sample were aged 3–14. By comparing the locations of small-city steel production in 1903 and 1930, I find that nearly all steel production that occurred in 1930 was in towns that also produced steel in 1903. Thus, individuals who were born in steel towns from 1916-1927 (and who remained in those towns through childhood) would have been exposed to pollution from birth through at least 1930. On the other hand, as mentioned above, many smaller steel production facilities shut down as the steel industry consolidated during the 1930s.[16]

The analysis below examines the relationship between later-life mortality and birth in a steel town. Growing up in a steel town could potentially affect old age health due to the long-reach consequences of pollution exposure, but it could have a negative impact on health for other reasons as well. For example, steel towns might differ from non-steel towns in terms of early-life disease exposure due to differences in geography (e.g., being on a river and population density). As a second example, the closure of steel production facilities surely led to job loss, and as Sullivan and von Wachter (2009) show, job displacement generally leads to increased mortality among men. In turn, this may have adversely affected the long-term health outcomes for steel workers' children.

Given these issues, one approach for trying to identify the impact of steel production *pollution* exposure on long-term health is to look at the elevation of towns. Many steel towns in Pennsylvania were in valley towns. At a minimum it is important to control for elevation as a way of being sure that any possible negative long-term health consequences to being born in a low-elevation city is not falsely attributed to pollution from steel production. Furthermore, pollution from steel production was likely to be particularly harmful in low-elevation valley towns, in which air inversions may have caused pollution to become trapped for extended periods.

A second approach focuses on the level of steel production in the town. An important advantage of 1930 data is that they include not only the location of steel production facilities, but also measures of the steel production capacity. This measure is the sum of 1000s of tons of steel production capacity for each plant in operation in the town. Thus I can also see if individuals who are born in steel towns with relatively high levels of production, and thus relatively high levels of pollution, also tend

[15]Unfortunately, it appears that Pennsylvania did not publish such records in the Annual Report of Industrial Statistics for years after 1903.

[16]Using American Iron and Steel Institute reports from 1954, for example, I find that virtually all of the small-town steel production facilities that operated in 1930 were no longer in operation as of 1954, when the individuals in my study sample were in their 20s and 30s.

to have disproportionately lower survival in old age.[17]

I also match the older-age residence zip code to the median income in the zip code to form a crude proxy of prosperity in later life. This variable serves as a control variable in many of my regression analyses below.

Table 1 provides a set of summary statistics for my sample: white individuals, with complete data records, for the birth cohorts 1916–1927 born in Pennsylvania. Among those who are excluded from the sample are individuals for whom it was not possible to match birth place to a "populated place" as given by the U.S. Geological Survey. Typically this happened if a county and city had the same name. The most important example is Philadelphia. Thus my starting sample excludes those individuals. In total the sample is quite large—over 780,000.

As noted in the introduction, I split the sample according to the size of birthplaces, my analysis focuses on people born in towns. Table 1 thus summarizes key variables according to the size of individuals' birthplaces. First are "Cities," corresponding to city sizes above the median (a birthplace reported to have over 2534 individuals in the SSA/Medicare data set). "Towns" are birthplaces with 2534 or fewer individuals (i.e., the median and below).

The first row on Table 1 shows that in larger cities, well over half of people were born in places with a steel mill. Indeed, steel production was especially concentrated in the largest cities; in 1930 there were steel mills in all but two of the 13 largest cities represented in this group (Wilkes-Barre and Hazelton). The second row shows that survival to age 75, conditional on being alive at 65, is approximately 80 percent and does not vary much across the size of one's birthplace. Finally, the lifetime income proxy—median zip code level income in the older-age residence—is somewhat higher for those born in larger cities than in smaller cities.

I conduct *within county* analyses of the association between mortality and birth in a steel town. The goal is to compare mortality for individuals who were born in steel towns to similarly-sized neighboring non-steel towns. As discussed in more detail below, very few counties have more than one large city, so within county analysis is viable only for towns. Thus, in Table 2, I focus on sample characteristics for the sample of more than 390,000 individuals born in these locations.

The first row of Table 2 shows that probability of surviving to age 75, conditional on being alive at age 65, is slightly lower for those born steel towns than in non-steel towns. As for other characteristics, a slightly higher fraction of those born in steel towns than non-steel towns reside

[17]In future work I also intend to pursue an additional strategy. At present I do not exploit the timing of the shutdown of steel production facilities across towns. By determining the dates of shutdowns it should be possible to form rough estimates of years of likely lifetime exposure for cohorts within towns, and this should further help with the empirical strategy.

in Pennsylvania in old age. Interestingly, people who migrate out of Pennsylvania tend to be more prosperous (according to the Income Proxy) than those who do not.[18] Both the means and standard deviations of the Income Proxy are very similar for those born in steel towns and non-steel towns. Finally, steel towns tend to have lower elevation than non-steel towns.

1.4.2 Research Design

The goal is to see if early-life exposure to high-polluting steel production is associated with older-age mortality. This entails conducting a multivariate analysis in which I compare individuals born in steel towns to individuals born in non-steel towns.

To highlight the challenges that lie ahead, consider the following regression model of survival:

$$S_i = \beta_0 + \beta_1 I_i + \sum_t \beta_t Z_i^t + \gamma X_i + \epsilon_i \tag{1.1}$$

where S_i is an outcome "survival" variable equal to 1 if individual i survives to age 75 and 0 if she or he dies (conditional on survival to age 65); I_i is 1 if the individual was born in a town with steel production in 1930 and 0 otherwise, and is meant to capture effects of early-life exposure to pollutants from steel production; Z_i^t is an indicator variable equal to 1 if the individual belongs to the specified birth cohort $\times$ gender cell (e.g., men born in 1927); and X_i is a vector of all other relevant factors that affect old-age survival.

Unfortunately, while the data include reasonably good measures for S_i, I_i, and Z_i^t, there are virtually no data for X_i. In regressions below, these omitted variables are subsumed in the error term, which of course is a problem if they are correlated with I_i.

To deal with this issue, in most of the analyses below I proceed by conducting *within county* comparisons using similarly-sized cities. The idea is that many of the early-life factors that might affect later-life health (i.e., variables in the vector X_i) are likely to be comparable in similarly-sized towns *within the same county.* To implement this idea, I include county fixed effects in most regressions reported below. There are 67 counties in Pennsylvania, of which 21 contain at least one steel town.

This strategy means that it is not especially credible to proceed with the "cities" in my analysis. The problem is that among the 67 counties in Pennsylvania, only seven have more than three cities. Beyond those counties, the presence of larger cities is very sparse within counties; there are four

[18] This means that if income and mortality are related, it might be that those who migrate out of Pennsylvania also have lower mortality. As noted below, an important advantage of the Duke/SSA Medicare data is that includes individuals who remain in their home towns and also those who move away.

counties with three cities, 11 counties with two cities, and 20 counties with only one city.[19]

Even for the small number of counties for which within-county comparisons among larger cities is technically feasible, it is probably not compelling. Consider, for example, Allegheny County. The largest city in that county is Pittsburgh. In 1930 Pittsburgh had a very large steel-production capacity, and many people born in Pittsburgh from 1916–1927 likely had high levels of air pollution exposure due to the steel production. However, while there were other "cities" in Allegheny County, it is difficult to argue that those cities would be otherwise comparable to Pittsburgh.[20] For this reason, I restrict all of the analysis below to *towns,* i.e., those with a population at or below the median in my data.

Beyond the use of county fixed effects, I have one additional control—the average income by zipcode at age 65 ("income proxy") variable. The inclusion of this variable is intended to help with the following issue: Suppose that steel production plants bring prosperity to a community, so that steel towns tend to have higher incomes. A large literature establishes a positive relationship between higher income levels and survival.[21] Below I find that inclusion of an income proxy variable does not alter the key results in the analyses; this provides some evidence that any correlation between town-level steel production and prosperity is not driving results.

With this research design in mind, again consider Table 2. Among the more than 390,000 individuals born in small towns, over 21,000 were born in steel towns. These individuals likely had on average much higher exposure to air pollutants when they were young than did those born in towns without steel production.

A key concern for the research design is that the data do not include the age at which individuals migrated out of their home town. Ideally, for the purposes of the study, migration rates would be very low at young ages—so that most of those born in steel towns do indeed get exposure to pollution at least throughout childhood and youth. While age of migration is not known (so duration of pollution exposure cannot be directly measured), it is possible to provide some indirect evidence on the issue. Figure 1 shows estimates of residence, by age, for individuals born in Pennsylvania,

[19] Even for counties with more than one city, there are cases where it is not possible to compare people born in steel and non-steel cities (conditional on being born in a city). To pick one example, Dauphin County has two cities, Harrisburg and Steelton. Both had steel production facilities in the early twentieth century.

[20] For instance, McKeesport and McKees Rocks are the two largest cities in Allegheny County aside from Pittsburgh (according to my population measure and based on population counts in 1920). While both cities are "cities" in my data, McKeesport has a population that is only approximately 1/9 as large as Pittsburgh, and McKees Rocks is only approximately 1/14 the size of Pittsburgh. (Making matters worse, both of those cities also had steel production facilities.)

[21] Grossman (1972) provides theoretical ideas on the topic. Empirical work shows strong associations between income and mortality, at the country level and at the individual level within countries. See, for example, Preston (2007), Preston and Elo (1995), Elo and Preston (1996), Cutler, *et al.* (2006), and Cutler and Lleras-Muney (2006, 2010).

1916-1927, calculated using 1920–2000 public-use Census samples. As it turns out, migration in early childhood was quite uncommon; fewer than 8 percent of Pennsylvanians migrated to another state by age 15 and only approximately 16 percent migrated by age 22. To the extent that these individuals did not get full early-life exposure to pollution, inclusion of these individuals in the sample leads to an *under-estimation* of the old-age mortality effects of pollution. As for those who remained in their home towns, pollution exposure would in all cases be through at least 1930, as noted above, and in a relatively small number of cases might have extended beyond 1954.[22] In the regression analyses below, samples include both those who remain in Pennsylvania and those who moved outside of Pennsylvania by old age.[23]

1.4.3 Results

The paper's first results are given in Table 3. Column (1) shows that in a regression in with a steel-town indicator variable and gender $\times$ cohort effects, there is a negative coefficient on Birth in a Steel City. People born in steel towns have lower old-age survival rates than those born in non-steel towns, and this association is statistically significant. Age $\times$ cohort effects (not reported in the Table) are as follows: The omitted category was men in birth cohort 1916. Then for men in other cohorts, estimated effects are generally quite small, with a bit of an upward drift for later cohorts. For women estimated cohort effects are very large, typically on the order of 0.09.[24]

Column (2) shows that inclusion of the income proxy does not much alter the key inference in column (1). As for the income proxy variable, as expected, the coefficient on this variable is positive and statistically significant; people who live in higher-income communities also live longer. More importantly, columns (3) and (4) show that the key inference is very similar for a specification that includes county fixed effects. This is a primary result in my analysis: *within counties,* people who were born in small steel-producing cities have lower survival rates than people born in small non-steel producing cities.[25]

To put this result into perspective, recall that the overall mortality rate for ages 65 to 75 is approximately 20 percent (see Table 1). Individuals born in steel towns have mortality that is

[22] In addition, some men born in steel towns might have moved to other steel-producing towns to get jobs in steel production.

[23] It is possible that migration out of Pennsylvania is selective. For instance, Halliday and Kimmitt (2008) show that in the U.S., people who are healthier tend to be more mobile. If I were to exclude migrants, this might lead to bias in my estimates of impact of pollution on older-age mortality.

[24] Among older individuals born in small town in Pennsylvania, 1916-1927, as in most populations, women survive at higher rates than do men.

[25] I estimated all regressions also including a population measure—the log of the number of people in the sample born in that location. This variable was not statistically significant, and there was virtually no change to other coefficients. In addition, I ran regressions in columns (1) and (3) using only those 21 counties which contained at least one town with steel production. Again, the coefficients do not change with this specification.

approximately 0.75 percentage points higher than comparable individuals born in non-steel towns. Thus mortality is approximately 4% higher for those born in steel towns.[26]

Columns (4) and (5) provide analyses separately for women and men. For both women and men, the coefficient on Born in Steel City is negative, though the coefficient is statistically significant only for women.

As mentioned above, it is possible that even within counties, steel cities and non-steel cities differ along unobserved dimensions that affect old-age mortality beyond any effects of air pollution. To give an example, suppose all steel towns were along major rivers while many non-steel towns were not, and suppose further that river towns were more susceptible to water-born diseases. Then it might be incorrect to attribute lower survival rates among those born in steel towns to pollution generated by the steel mills. To address that problem, I took the sample of steel towns and divided them into towns that in 1930 had steel production capacity that was lower than the median and higher than the median (using data from American Iron and Steel Institute, 1930). Results, presented in Table 4, suggest that low survival rates for individuals born in steel cities is not due to being born in a steel city *per se* but instead is the consequence of *relatively high levels of steel production.*[27] The results are consistent with the idea that reduced old-age survival associated with exposure to steel production is due primarily to high levels of exposure. Mortality from ages 65 to 75 is approximately one percentage higher for those born in steel cities with above-median steel production capacity than for those born in non-steel towns within the same county.

Columns (2) and (3) of Table 4 show similar patterns for women and men. This is interesting because, it may be presumed, some men born in steel towns may have worked as steel workers in adulthood, in which case being born in a steel city might affect older-age health via any lasting effects of working in that industry. The same is much less likely for women of this era. This evidence is consistent with an interpretation that attributes a role for early-life exposure to pollution for generating excess mortality among people born in steel towns.

The final piece of evidence concerns a potential role for elevation in shaping my findings. Among individuals born in mill towns, it seems likely that exposure to PM would have been worse for those born in low-elevation steel towns, i.e., towns that were in valleys, which would have typically accumulated higher concentrations of PM due to atmospheric inversions. A simple way of evaluating that idea is to see if there is a correlation between survival probabilities and the elevation of the mill

[26]This analysis was replicated using an alternate definition of steel-town indicator—towns are designated to be steel towns only if they appear in both of two historical sources, Secretary of Internal Affairs (1903) and American Iron and Steel Institute (1930). Results are virtually unchanged when I use that definition.

[27]Note that capacity, i.e., amount of steel produced, is proportional to the amount of coal input.

town (which was determined using Geographical Analysis Tools, 2012).

In Table 5, I report the results of an analysis that incorporates elevation. First I take the basic regressions from Table 4 but add "Elevation" (measured in 1000s of meters). Results are reported in column (1). Elevation does not have a statistically significant impact on survival in the regression. The specification reported in column (2) also includes two interaction terms: first, an interaction between Elevation and "Born in Town with Below-Median Steel Production Capacity" and, second, an interaction between Elevation and "Born in a Town with Above-Median Steel Production Capacity." Finally, column (3) reports this same specification but with county fixed effects. In this last regression, the main effects of being born in either type of steel town are negative, and the coefficient on "Born in Town with Above-Median Capacity" is sizable (-0.025) and is highly statistically significant. Importantly, the main effect of "Elevation" is close to 0. This suggests that any relationship between birth in a steel town and old-age survival is *not* due to an omission of elevation in my regressions. Coefficients on the interaction terms are positive, but neither is statistically significant in the specification with county fixed effects.

Figure 2 illustrates the estimated relationship for those born in small steel cities with above-median production using the coefficients estimated in the specification with county fixed effects, column (3).[28] For those born in towns near sea level (elevation 0), survival is -0.025 lower than in the reference group (small cities with no steel production).[29] The negative impact on survival associated with being born in an above-median capacity steel steel town becomes more moderate at higher elevations. To put this in perspective, note that the estimated impact on survival of being born in a steel town with above-median capacity is 0 at an elevation of approximately 380 meters, as $-0.025 + (0.059 \times 0.380) \approx 0$. In my sample, among those born in towns with above-median steel production, more than 98% were born in places with elevation below 380 meters.

1.5 Conclusion and Directions for Future Research

Looking at a sample of over 390,000 individuals born in small cities in Pennsylvania, 1916–1927, I find that individuals who were born in steel towns have significantly higher rates of mortality post-age 65 than those born in comparable towns that did not have steel production facilities. There are three potentially important features of this association. First, the relationship holds for people born in neighboring towns, i.e., within the same county. Second, the relationship between old-age

[28]The coefficient on the interaction of elevation and being born in a town with above-median capacity is statistically significant in the specification without county fixed effects, but not the one with the county fixed effects. I nonetheless focus on the specification with county fixed effects since my research design relies on the fixed-effects approach.

[29]In fact, the lowest elevation points in Pennsylvania are 0. These are locations along the Delaware River.

mortality and birth in a steel town is stronger in towns that had relatively higher levels of steel production. Third, there is some evidence that old-age mortality is especially high for individuals born in places with relatively high levels of steel production and relatively low elevation. This last finding is consistent with the possibility that low-elevation locations were subject to atmospheric inversions that tended to trap air pollution, thereby increasing pollution exposure.

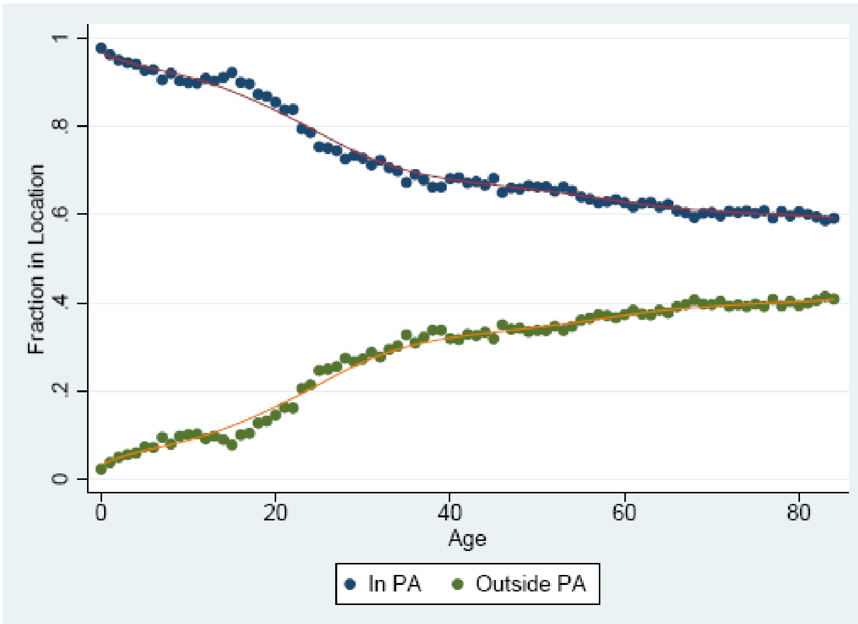
This evidence is consistent with an important idea in the literature—that exposure to air pollution at young ages can lead to latent disease processes that affect morbidity and mortality later in life. To my knowledge, my paper is the first to provide evidence that growing up in high-pollution town is associated with higher mortality in old age. This finding is a potentially useful contribution because most of the available evidence on health impacts of air pollution focus on contemporaneous relationships; very little work examines the "long reach" effects of pollution on health.

Of course, caution should be exercised in interpreting the results reported here. It could be that the adverse health impacts associated with birth in a steel town are not due to air pollution, but rather to other health threats that are especially prevalent in steel towns. To give just one example, infectious disease exposure might have been higher in steel towns. It is important that future work continues to explore possible mechanisms that lead the observed associations between place of birth and older-age mortality.[30]

There are several important directions for future work. For example, having access to data which indicates timing of plant closures would be helpful for determining length of exposure. Perhaps the most important additional analysis would focus on "cause of death;" it would be extremely valuable to see if there are differences in patterns of disease processes for individuals who grew up in steel towns compared to those who did not. For future research, I am hopeful that for a subset of my sample, records can be matched to cause of death or other data on health outcomes from Medicare records.

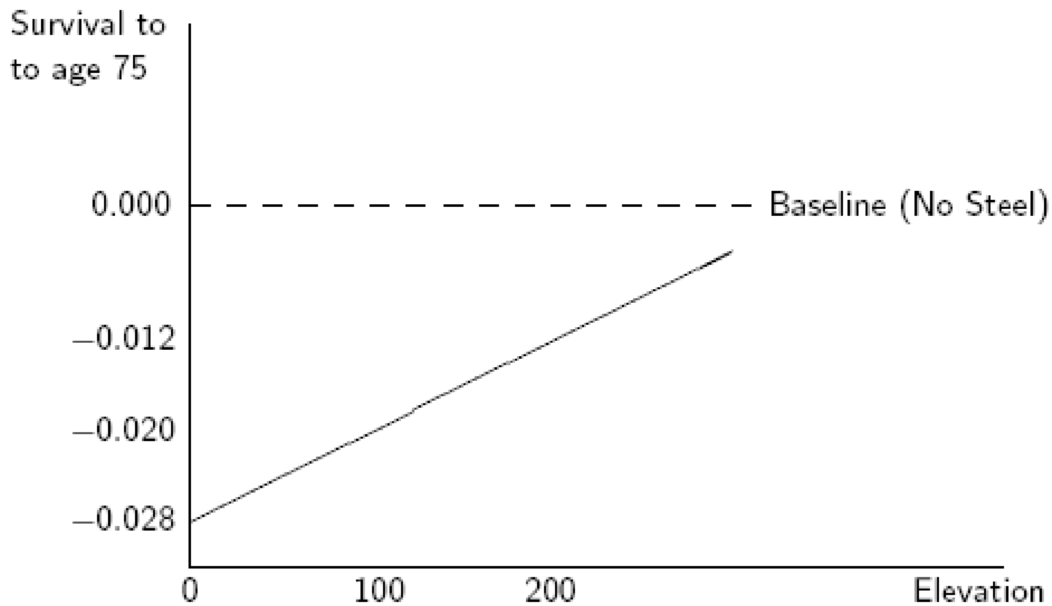
[30]On infectious disease specifically, I did check to see if the cohort born during the 1918 influenza pandemic had a disproportionate impact on results. (See, e.g., Almond, 2006, who shows that being born in that year may have had lasting detrimental health impacts.) Regression are very similar when that cohort is dropped from the sample. Also, it is worth noting that if nearby towns were subject to similar infectious disease exposure, then the within-county design would help deal with this issue. Still, further data collection and analysis is surely important.

Figure 1.1: Fraction of the Population In Pennsylvania and Outside of Pennsylvania by Age, Cohorts Born 1916-1927



Source: Calculations from U.S. Census data, public-use samples, 1930–2000.

Figure 1.2: Estimated Relationship Between Elevation and Survival, Individuals Born in Towns with Above-Median Steel Production



Source: Graph based on estimated coefficients from regression (1) in Table 5.

Table 1.1: Sample Characteristics, Individuals Born in Pennsylvania, 1916–1927

	(1) Cities (Above Median)	(2) Towns (Below Median)
Proportion Born in Steel City	0.555	0.055
	(0.497)	(0.228)
Prop. Surviving to Age 75 Given Survival to 65	0.800	0.802
	(0.400)	(0.398)
Income Proxy (in $1000s)	42.3	41.7
	(14.9)	(13.4)
Number of Municipalities	92	5,520
Sample Size	389,747	391,851

Note: Author's calculations using the Duke SSA/Medicare Data Set. "Cities" are birthplaces with more than 2534 individuals in the data (which is the median) and "Towns" are birthplaces with with 2534 or fewer individuals in the data (the median and below). The income proxy variable is the median annual income (in 2000 dollars) in the zip code where the individual resides post age 65. Statistics are presented for a sample that includes whites only, for those for whom data are complete. The sample excludes Philadelphia (see text). Standard deviations are in parentheses.

Table 1.2: Sample Characteristics for Individuals Born in Pennsylvania Towns, 1916-1927

	(1) Steel Towns	(2) Non-Steel Towns
Proproption Surviving to Age 75	0.795	0.802
Given Survival to 65	(0.404)	(0.400)
Proportion Residing in PA	0.684	0.666
post Age 65	(0.465)	(0.471)
Income Proxy (in $1000s)		
Residing in PA post Age 65	38.7	39.4
	(10.6)	(11.1)
Not Residing in PA post Age 65	46.2	46.2
	(16.4)	(16.4)
All	41.1	41.7
	(13.2)	(13.5)
1930 Steel Production Capacity	362.4	–
	(1062.6)	
Elevation of Birthplace (Meters)	241.1	284.7
	(86.9)	(135.8)
Number of Towns	45	4575
Sample Size	21,459	370,392

Note: Author's calculations using the Duke SSA/Medicare Data Set matched to historical records on the location of steel production. "Towns" are as defined in Table 1. Steel Production Capacity is measured in 1000s of tons per years (as of 1930).

Table 1.3: Survival to Age 75 and Birth in a Steel Town, Individuals Born in Pennsylvania Towns, 1916–1927

	(1) All	(2) All	(3) All	(4) All	(5) Women	(6) Men
Born in Steel Town	-0.0075^{**}	-0.0068^{**}	-0.0087^{***}	-0.0074^{**}	-0.0089^{**}	-0.0054
	(0.0030)	(0.0030)	(0.0031)	(0.0031)	(0.0037)	(0.0052)
Income Proxy	–	0.0012^{***}	–	0.0012^{***}	0.0011^{***}	0.0014^{***}
		(0.000049)		(0.00049)	(0.000060)	(0.000083)
Cohort × Gender F.E.?	yes	yes	yes	yes	–	–
Cohort F.E.?	–	–	–	–	yes	yes
County F.E.	–	–	yes	yes	yes	yes
Sample Size	391,851	391,851	391,851	391,851	220,176	171,675

Note: Author's calculations using the Duke SSA/Medicare Data Set matched to historical records on the location of steel production. Dependent Variable is "Survival to Age 75" is conditional on survival at age 65. The sample includes only those who are born in populated places in Pennsylvania, in towns (see Table 1 for definition). There are 67 counties in Pennsylvania, of which 21 contain at least one steel town. Standard errors, clustered at the town level for (1) and (2), are in parentheses. Significance: *0.10, **0.05, ***0.01.

Table 1.4: Survival to Age 75 and Birth in Towns with Low Steel Production Capacity and High Steel Production Capacity

	(1) All	(2) Women	(3) Men
Born in Town with Below-Median Capacity	-0.0022	-0.00071	-0.0043
	(0.0033)	(0.0040)	(0.0054)
Born in Town with Above-Median Capacity	-0.011^{***}	-0.011^{***}	-0.010^{*}
	(0.0033)	(0.0040)	(0.0055)
Income Proxy	0.0012^{***}	0.0011^{***}	0.0014^{***}
	(0.000049)	(0.000060)	(0.000083)
Cohort × Gender F.E.?	yes	–	–
Cohort F.E.?	–	yes	yes
County F.E.?	yes	yes	yes
Sample Size	391,851	220,176	171,675

Note: Author's calculations using the Duke SSA/Medicare Data Set matched to historical records on the location of steel production. Dependent Variable is "Survival to Age 75" is conditional on survival at age 65. The sample includes only those who are born in populated places in Pennsylvania, in towns (see Table 1 for definition). Standard errors, clustered at the town level for (1), are in parentheses. Significance: $^{*}0.10$, $^{**}0.05$, $^{***}0.01$.

Table 1.5: Survival to Age 75, Birth in a Steel Town and the Role of Elevation

	(1)	(2)	(3)
Born in Town with Below-Median Capacity	0.00096	-0.012	-0.0061
	(0.0036)	(0.0088)	(0.0087)
Born in Town with Above-Median Capacity	-0.0091^{***}	-0.028^{***}	-0.025^{***}
	(0.0028)	(0.010)	(0.0095)
Elevation (Kilometers)	0.0073	0.0048	-0.0019
	(0.0059)	(0.0060)	(0.0094)
Elev. $\times$ Born in Town with Below-Median Capacity	–	0.063	0.018
		(0.043)	(0.039)
Elev. $\times$ Born in Town with Above-Median Capacity	–	0.079^{**}	0.059
		(0.040)	(0.037)
Income Proxy	0.0012^{***}	0.0012^{***}	0.0012^{***}
	(0.000049)	(0.000049)	(0.000049)
Cohort $\times$ Gender Effects?	yes	yes	yes
County Fixed Effects?	–	–	yes
Sample Size	391,851	391,851	391,851

Note: Author's calculations using the Duke SSA/Medicare Data Set matched to historical records on the location of steel production. Dependent Variable is "Survival to Age 75" is conditional on survival at age 65. Elevation is in 1000s of meters. The sample includes only those who are born in populated places in Pennsylvania, in small cities (see Table 1 for definition). Standard errors, clustered at the town level for (1), are in parentheses. Significance: *0.10, **0.05, ***0.01.

References

American Iron and Steel Institute. 1930. *Directory of the Iron and Steel Works of the United States and Canada, 21st Edition.* Baltimore: The Lord Baltimore Press.

American Iron and Steel Institute. 1954. *Directory of the Iron and Steel Works of the United States and Canada, 21st Edition.* Baltimore: The Lord Baltimore Press.

Almond, Douglas. 2006. "Is the 1918 Influenza Pandemic Over? Long-Term Effects of *In Utero* Influenza Exposure in the Post-1940 Population," *Journal of Political Economy,* 114(4), 672-712.

Almond, Douglas, Lena Edlund, and Mrten Palme. 2009. "Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden," *Quarterly Journal of Economics,* 124(4), 1729-1772.

Barker, D. J. P. 1995. "Fetal Origins of Coronary Heart Disease," *British Medical Journal,* 311, 171-74.

Batchelor, R. P. *et al.* 1926. "A Clinical and Laboratory Investigation of the Effect of Metallic Zinc, of Zinc Oxide, and of Zinc Sulphide Upon the Health of Workmen." *Journal of Industrial Hygiene* 8, 322-363.

Beer, Worrell, and Blok. 1998. "Future Technologies for Energy-Efficient Iron and Steel Making." *Annual Review of Energy Environment*, 23, 123-205.

Billings, Paul. 2011. "Emissions of Hazardous Air Pollutions From Coal-Fired Power Plants." *Environmental Health and Engineering,* EH&E Report 17505.

Black, Dan, *et. al.* 2012 "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South," manuscript, University of Chicago.

Bleiwas, Donald and DiFrancesco, Carl. 2010. "Historical Zinc Smelting in New Jersey, Pennsylvania, Virginia, West Virginia, and Washington, D.C. with Estimates of Atmospheric Zinc

Emissions and Other Materials." U.S. Department of the Interior and U.S. Geological Survey, Report 2010-1131.

Buchauer, M. J., 1973. "Contamination of Soil and Vegetation Near a Zinc Smelter by Zinc, Cadmium, Copper, and Lead." Environmental Science and Technology. 7, 131-135.

Carroll, M. D. and Loneragan, J. F. 1968. "Response of Plant Species to Concentration of Zinc in Solution: I. Growth and Zinc Content of Plants." *Aust. J. Agric. Res.* 19, 859-868.

Case, Anne, Angela Fertig, and Christina Paxson, 2005. "The Lasting Impact of Childhood Health and Circumstance," *Journal of Health Economics,* 24(2), 365-389.

Case, Anne, Darren Lubotsky, Christina Paxson, 2002. "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review,* 92(5), 1308-1334.

Chay, Ken, Carlos Dobkin and Michael Greenstone. 2003. "The Clean Air Act of 1970 and Adult Mortality." *Journal of Risk and Uncertainty.*

Chay, Ken and Michael Greenstone. 2003a. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *Quarterly Journal of Economics.*

Chay and Greenstone. 2003b. "Air Quality, Infant Mortality, and the Clean Air Act of 1970." Center for Energy and Environmental Policy Research.

Chmielewski, J., Jaremin, B., Martnicki, C, Konieczka, R. 1947a. "Evaluation of Occupational Exposure to Zinc Oxide in the Marine Production-Shipyard: I. Examination of the Working Environment and the Stands Under Exposure." *Bull. Inst. Mar. Med. Gdansk,* 25, 43-51.

Chmielewski, J., Jaremin, B., Martnicki, C, Konieczka, R. 1947b. "Evaluation of Occupational Exposure to Zinc Oxide in the Marine Production-Shipyard: II. Examination of the State of Health of the Workers Exposed to Zinc Oxide." Bull. Inst. Mar. Med. Gdansk 25, 53-85.

Climate Leaders. 2003. “Direct Emissions from Iron and Steel Production.” U.S. Environmental Protection Agency.

Currie, Janet and Reed Walker. 2011. “Traffic Congestion and Infant Health: Evidence from E-ZPass.” *American Economic Journal: Applied Economics.* 3(1), 65-90.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney, 2006. “The Determinants of Mortality,” *Journal of Economic Perspectives,* 20(3), 97-102.

Cutler, David, and Adriana Lleras-Muney, 2010. “Understanding Differences in Health Behaviors by Education ,” *Journal of Health Economics,* 29(1), 1-28.

Davis, Devra. 2002. *When Smoke Ran Like Water.* New York: Basic Books.

DiFrancesco, Kelly, Bleiwas, and Fenton. 2010. “Iron and Steel Statistics.” U.S. Geological Survey.

Dockery, Douglas W., *et al.* 1993. “An Association Between Air Pollution and Mortality in Six U.S. Cities.” *The New England Journal of Medicine,* 329(24), 1753-59.

Durand, Michael, et al. 2004. “Elevated Trace Element Output in Urine Following Acute Volcanic Gas Exposure.” *Journal of Volcanology and Geothermal Research,* 134(1-2) 139-148.

Elo, Irma T., and Samuel H. Preston, 1996. “Educational Differentials in Mortality: United States, 1979-1985,” *Social Science & Medicine,* 42(1), 47-57.

Environmental Protection Agency. 1987. “Summary Review of Health Effects Associated with Zinc and Zinc Oxide.” Health Issue Assessment, EPA Report 600/8-87/022F.

Fierro, M. 2000. “Particulate Matter.” Pima County Department of Environmental Quality Air.

Fierro, M., O’Rourke, M.K., and Burgess, J.L. 2001. “Adverse Health Effects of Exposure to Ambient Carbon Monoxide.” Prepared for the University of Arizona College of Public Health.

Finkelstein, Jerrett, and Sears. 2004. "Traffic Air Pollution and Mortality Rate Advancement Periods." American Journal of Epidemiology.

Fogel, Robert, 2004. *The Escape from Hunger and Premature Death, 1700-2100.* Cambridge University Press.

Gasaway, W. C. and Buss, I. O. 1972. "Zinc Toxicity in Mallard Duck." *Journal of Wildlife Management.* 36, 1107-1117.

Gauderman, W. J., *et al.* 2004. "The Effect of Air Pollution on Lung Development for 10 to 18 Years of Age," *New England Journal of Medicine,* 351(1), 1057-1067.

Gauderman, W. J., *et al.* 2007. "Effects of Exposure to Traffic on Lung Development from 10 to 18 Years of Age: A Cohort Study," *The Lancet,* 369(9561), 571-577.

Geographical Analysis Tools accessed from veloroutes.org, 05/07/2012.

Grossman, Michael, 1972. "On the Concept of Health Capital and the Demand for Health," *The Journal of Political Economy,* 80(2), 223-255.

Halliday, Timothy, and Michael Kimmitt, 2008. "Selective Migration and Health in the USA, 1984-93," *Population Studies,* 62(3), 321-334.

Hayward, Mark D., and Bridget K. Gorman, 2004. "The Long Arm of Childhood: The Influence of Early-Life Social Conditions on Men's Mortality," *Demography,* 41(1), 87-107.

Jacobson, Mark. 2008. "On the causal link between carbon dioxide and air pollution mortality." *Geophysical Research Letters,* 35.

Lam, H. F., Conner, M. W., Rogers, A. E., Fitzgerald, S., and Amdur, M. O. 1985. "Functional and Morphologic Changes in Lungs of Guinea Pigs Exposed to Freshly Generated Ultrafine Zinc Oxide," *Toxical. Appl. Pharmacology,* 78, 29-38.

Lave, Lester B. and Freeburg, Linnea C. 1973. "Health Effects of Electricity Generation from Coal, Oil, and Nuclear Fuel." Tepper School of Business. Paper 1173.

Maccini and Yang. 2009. "Under the Weather: Health, Schooling, and Economic Consequences of Earli-Life Rainfall." *American Economic Review.*

Pope, C. A. and D. W. Dockery. 2006. "Health Effects of Fine Particulate Air Pollution: Lines that Connect." *Journal of the Air and Waste Management Association.* 56, 709-742.

Pope C. A. 1989. "Respiratory Disease Associated with Community Air Pollution and a Steel Mill, Utah Valley." *American Journal of Public Health.*

Pope, C. A., *et al.* 1992."Daily Mortality and PM_{10} Pollution in Utah Valley." *Archives of Environmental Health: An International Journal,* 47(3).

Pope, C. A., *et al.* 1995. "Particulate Air Pollution as a Predictor of Mortality in a Prospective Study of U.S. Adults," *American Journal of Respiratory and Critical Care Medicine.* 151(3), 669-674.

Preston, Samuel H., 2007. "The Changing Relation between Mortality and Level of Economic Development," *International Journal of Epidemiology*, 36(3), 484-490.

Preston, Samuel H., and Irma T. Elo, 1995. "Are Educational Differentials in Adult Mortality Increasing in the United States?" *Journal of Aging and Health,* 7(4), 476-496.

Radian Corporation. 1989. "Health Hazard Assessment Summary: Steel Mill Emissions." United States Environmental Protection Agency, Report 450/3-90-026.

Secretary of Internal Affairs of the Commonwealth of Pennsylvania. 1903. "Annual Report Part III: Industrial Statistics." Wm. Stanley Ray, State Printer of Pennsylvania.

Somers, Christopher M., Brian E. McCarry, Farideh Malek, and James S. Quinn. 2004. "Reduction of particulate air pollution lowers the risk of heritable mutations in mice," *Science,* 304(5673), 1008-1010.

Sullivan, Daniel, and Von Wachter, Till. 2009. "Job Displacement and Mortality: An Analysis using Administrative Data," *Quarterly Journal of Economics,* 124(3), 1265-1306.

Warren, Kenneth. 1973. *The American Steel Industry 1850-1970: A Geographical Interpretation.* Oxford: Clarendon Press.

Warren, Kenneth. 2001. *Wealth, Waste, and Alienation: Growth and Decline in the Connellsville Coke Industry.* Pittsburgh: Pittsburgh University Press.

Wellenius, G. A., *et al.* 2003. "Inhalation of Concentrated Ambient Air Particles Exacerbates Myocardial Ischemia in Conscious Dogs," *Environmental Health Perspectives,* 111(4), 402-408.

CHAPTER II

“Tell All the Truth, but Tell it Slant”: Testing Models of Media Bias [1]

2.1 Introduction

Social and behavioral scientists across several disciplines, including economics, psychology, and political science, are interested in the causes and consequences of media bias. News content can have significant effects on beliefs, actions and outcomes. However, it can often be difficult to assess the existence and impact of bias. First, most measures of bias are by necessity comparative. It can be difficult to ascertain whether the *New York Times'* political reporting is liberally biased relative to the truth, but much easier to ascertain whether it is liberally biased relative to Fox News. Second, bias may be caused by individuals' beliefs about likely outcomes, but biased media reports can also lead to changes in outcomes. For example, if the media is biased towards reporting an Obama victory, this may in fact make an Obama victory more likely. Sorting out these complications is a potentially important undertaking; many people believe that media bias can have significant welfare effects, causing rising polarization and mistrust in the media.

Of course, addressing if, and to what extent, consumers are harmed by bias requires a theory of media bias as well as an empirical assessment of key predictions generated by the theory. Moreover, many models of media bias can predict a variety of types of bias depending on the parameters of the models, and so it is necessary to test not just the direction of bias but also how bias changes with the environment. Therefore it is important to find data where there is exogenous variation in the environment in order to test the comparative static predictions of models of media bias.

[1] The first line of the title is due to a poem of Emily Dickinson, who incidentally often wrote in slant verse. We would like to thank Yan Chen, Stefano DellaVigna, Rachel Kranton, Yusufcan Masatlioglu, Stephen Salant, Daniel Silverman, Mel Stephens, and Lowell Taylor for helpful comments, as well as seminar participants at the Midwest Economics Association Annual Meeting (2012) and Oxford. Any and all errors are ours alone. Emails: collin.raymond@economics.ox.ac.uk and setay@umich.edu

In this paper, we develop models of news provision and test predictions using novel data, which allow us to identify possible driving mechanisms behind media bias. Such analysis gives further insight into the welfare impact of bias. We document bias in a very simple setting — historical weather reports by the *New York Times*. In the late 1800s Manhattan was home to a professional baseball team, the New York Giants. Before June 16, 1896 the *New York Times* used government-produced weather predictions. On June 16, 1896 the *New York Times* switched to producing in-house weather reports. Our focus is the time frame, June 16, 1896 through the end of the century. For this period, we find an interesting form of bias: the *Times* was more accurate at predicting sunny weather and less accurate at predicting rainy weather on days when the Giants played home games, relative to non-home game days.

Our research design has several benefits. First, it features a regime shift in the weather reporting by the *New York Times* from the use of Weather Bureau reports to in-house weather reporting. Second, there is a clear exogeneity in realized outcomes — changes in the *Times'* weather predictions do not cause the realized weather to change. Third, we can observe both the predictions and the realized weather, meaning we can develop a measure of absolute bias, not simply comparative bias. Fourth, the type of bias we observe should lead to different posteriors on the part of the consumers, if they are aware of it. This is in contrast to many other papers, which study how different media outlets report the same event using different terms, leading to difficulty in assessing how individual beliefs should respond to biased information. Fifth, we can easily observe how bias changes with the parameters of the model. Last, but not least, the actual environment closely aligns with a tractable model.

A theory of media bias that encompasses all the important facets of the phenomena would be quite extensive. In this paper we set for ourselves a much simpler goal: We seek to identify possible driving mechanisms behind media bias in a stylized setting in which the media provides predictions about two potential states of the weather ("rainy" or "fair"). Given preferences, and constraints on how accurate weather reports can be, we then generate predictions and comparative statics regarding consumer-induced preferences for accuracies. In turn, we develop equilibrium predictions regarding how news-providing firms will respond to such consumer preference — in particular whether firms will provide differently biased news when consumer's actions, priors and payoffs change. We show how the environment and preferences can interact to generate biased signals; different mechanisms yield different predictions regarding the observed bias in news reports.

As we have noted, we test our predictions by evaluating the daily weather reports in the *New York Times* during the late nineteenth century. In the 1890s the *New York Times* published a daily weather report in its morning edition which was extremely simple, typically offering either a rather straightforward prediction for fair weather or for rain. This makes the reports easy to interpret for the purpose of statistical analysis. From 1890 through mid-1986 the *New York Times* used weather reports produced by the U.S. Weather Bureau in Washington D.C., and its predecessor, the Signal Service.[2] These forecasts were not very accurate (see Nekeber, 1995), but they were produced by a trusted independent source; it seems unlikely that they were slanted so as to please New York baseball fans. On June 16, 1896, the *New York Times* began producing the daily forecast in-house.[3] Since weather reports continued to be inaccurate (as meteorological technology was in its infancy), forecasters could easily accommodate a moderate amount of bias without greatly reducing overall accuracy.

The baseline test is to see if the *Times* was more likely to provide biased weather predictions on days when the New York Giants were scheduled to play a baseball game in their home park, the Manhattan Polo Grounds. We look for bias in the *Times* weather reports on home game days after the switch to in-house forecasting. We observe that reports are indeed biased for the subsample of home game days after June 16, 1896, and that correct reports for fair weather increase while correct reports of rainy weather decrease, i.e., weather reports are "optimistic" about the probability of fair weather on home game days.

We run additional tests to understand how the bias changes as the environment changes — in particular what occurs when the priors held by consumers or the payoff of attending a game change. We find that as months have a higher ex-ante probability of rain, the amount of bias grows. Noting that the Giants held various positions in the National League rankings in the years after the switch to in-house forecasting, we also examine how the bias changes with the yearly ranking of the baseball team. It appears that the amount of bias grows as the baseball team does better. These comparative static results are inconsistent with the predictions of our rational models of media bias. In particular, we find that the evidence does not support the bias being driven by consumers' desire to better match their action to the state, by the *New York Times* trying to increase attendance at Giants' games, nor by reputational concerns (as in Gentzkow and Shapiro, 2006).

Our paper contributes to a burgeoning literature in economics examining media bias. There are several existing models of media bias, including Mullainathan and Shleifer (2005), Suen (2004),

[2]Newspaper forecasts were the primary avenue by which people learned about the weather, since radio did not become popular until decades later. In fact, the first radio broadcast for entertainment purposes was not transmitted until 1919.

[3]Evidence for such change comes from the fact that the *New York Times* always cited the Weather Bureau before June 16, 1896. On and after this date, the *Times* did not cite the Weather Bureau. As an additional point, during the period of in-house forecasting, the *New York Times* predictions diverged from the *Wall Street Journal* (as we discuss in more detail below), which sometimes cites the Weather Bureau as a source.

Gentzkow and Shapiro (2006), and Baron (2006). In Mullainathan and Shleifer (2005) bias in news reporting arises because readers have a form of confirmation bias; they generally hold views that differ from the true state of the world and get disutility from reading news that is inconsistent with those views. Gentzkow and Shapiro (2006) present a model of news bias with Bayesian consumers who value accuracy in news, and infer the accuracy of reports based on priors. In this setting, news outlets bias toward reader priors as a way of boosting inferred quality. Gentzkow and Shapiro (2008) discuss various mechanisms for bias and show how competition can either enhance or ameliorate bias. Our theoretical model is also closely related to Gentzkow and Kamenica (2011), who look at how a principal can persuade Bayesian agents by utilizing different information structures.

There has also been a great interest in identifying bias in the media. Many papers have focused on determining the extent and causes of media bias. Most of these papers, including Groseclose and Milyo (2005), Gentzkow and Shapiro (2004, 2010, 2011), Gentzkow, Shapiro and Sinkinson (2012), Gentzkow, Petek, Shapiro and Sinkinson (2012), and Durante and Knight (2012) focus on the causes of bias in the context of political ideology. In contrast, DellaVigna and Kaplan (2007) and Gentzkow, Shapiro and Sinkinson (2012) examine the impact of media bias on behavior, again in a political context. DellaVigna and Gentzkow (2010) survey the evidence on persuasion in markets, and consider different forces that could drive persuasion. In a related vein of work, Oster, et al. (2011) also provides a behavioral framework for understanding the motivation behind preference for news.

The rest of the paper is organized as follows. Section 2 develops a model of the market for news and presents the implications for consumer demand and equilibrium provision of information. Section 3 details the data and presents results of the empirical tests. Section 4 concludes.

2.2 Models of Information Provision

In this section we present two possible models of information provision. Our setting is stylized; we consider a world with two payoff relevant states and binary signals. Individuals have two actions — a safe action, where the payoff is invariant to the state, and a risky action, in which the payoff depends on the state. Despite this simplicity, we believe it captures the essential details of the empirical work that follows.

In line with the literature, we distinguish between multiple sources of bias; we consider the implications of "supply driven" and "demand driven" bias. If bias is supply driven, firms receive a payoff that varies with the decision maker's action. If bias is demand driven, consumers have preferences over informational structures, and a profit maximizing firm provides signals in line with those preferences. We focus on a setting with a monopolistic provider of information.[4] The firm chooses the information structure to provide, but does not vary the price once a structure is chosen (similar to the fact that we do not observe variation across game days and non-game days in the price of the *New York Times*). Below, we consider the predictions of models where individuals behave according to expected utility, i.e., we consider "neoclassical" models. In our conclusion we discuss the implications of models where consumers have intrinsic preferences over information.

2.2.1 Environment

We assume that there are two states of the world, A (good weather) and B (bad weather). A decision maker takes one of two actions a (attend a game) and b (stay at home). Payoffs to the decision maker depend on action i and state j, and are denoted $u(i, j)$. Action a is the risky action, $u(a, A) > u(a, B)$, while b is safe, $u(b, A) = u(b, B)$.[5] Neither action dominates, and so it is better to match action to state: $u(a, A) > u(b, A) > u(a, B)$. We normalize the payoff of staying at home $u(b, A) = u(b, B) = 0$, and denote the relatively high payoff of going to the game in good weather $u(a, A) = H > 0$ and the lower payoff of going to the game in bad weather as $u(a, B) = L < 0$.

There is also a news provider which has access to a set of prediction technologies. We characterize a prediction technology as an information structure with two accuracies: If it will be good weather, it generates a signal that indicates *fair* with probability p. If the day will be bad weather, it indicates *rainy* with probability q. Throughout the paper we will assume that both firms and consumers know the chosen technology's values of p and q. In the case of demand-driven bias, since firms are responding to consumer's preferences, consumers have to realize at some level that p and q are changing. Otherwise, firms have no incentive to change p and q. In the case of supply-driven bias, the comparative statics will be exactly the same if consumers are naive and do not realize that the firm is manipulating p and q.

Given a prior $0 < \rho < 1$ on state A (fair weather), along with a particular information structure (p, q), we can easily find a pair of posteriors for a decision maker. After observing a fair signal the posterior for A is

$$\psi_F = \frac{\rho p}{\rho p + (1-\rho)(1-q)}. \tag{2.1}$$

[4]Focusing on a single firm makes the model more transparent. Moreover, the monopolistic information provided in our model of demand-driven bias provides the consumer an optimal information structure, a finding that would be replicated under competition. Therefore the comparative statics will be equivalent. In the case of supply-driven bias, if all firms receive a benefit from consumers attending the game, then again we would expect to see the exact same comparative statics. If on the other hand some firms do not receive a benefit from consumer actions, then we would not expect to see supply-driven bias in the marketplace at all.

[5]The results below can easily be generalized to accommodate the case where action b is also risky, so long as the difference in payoffs across states given action b is smaller than the difference in payoffs across states given action a.

After observing a rainy signal the posterior is

$$\psi_R = \frac{\rho(1-p)}{\rho(1-p)+(1-\rho)q}. \tag{2.2}$$

We structure our model so that a fair weather prediction increases beliefs about the weather being good, while a bad weather prediction decreases them. This implies that $p+q \geq 1$, which is without loss of generality within the space of all signals, $(p,q) \in [0,1] \times [0,1]$. We show these results, and provide an additional observation in the following lemma.[6]

Lemma 1 *The set $\{(p,q)|(p,q) \in [0,1] \times [0,1]$ and $p+q \geq 1\}$ satisfies three properties:*

1. *Observing a fair signal increases the posterior on state A relative to the prior, and observing a rainy signal decreases the posterior on state A relative to the prior.*

2. *For any signal structure $(p',q') \in [0,1] \times [0,1]$, there exists a $(p,q) \in \{(p,q)|(p,q) \in [0,1] \times [0,1]$ and $p+q \geq 1\}$ that generates the same posteriors with the same probabilities as (p',q').*

3. *For any strict subset S of $\{(p,q)|(p,q) \in [0,1] \times [0,1]$ and $p+q \geq 1\}$, there exists a point $(p',q') \in [0,1] \times [0,1]$ such that there is no element of S that generates the same posteriors as (p',q').*

The first point of the lemma shows that our convention of considering only $p+q \geq 1$ leads to the desired natural interpretation of the "accuracy" probabilities, p and q. The second point establishes our claim about generality. The third point provides a helpful backdrop as we turn to our specification of weather prediction technology.

Our approach to the prediction technology is to assume an inherent tradeoff between the accuracy of predicting fair weather, p, and the accuracy of predicting rain, q. In order to increase accuracy in one state, a news provider must reduce it in the other. To model this, we define a convex set of feasible signals $\Phi(p,q)$.[7] We define the production frontier of the feasible set using $\phi(p)$ — the maximal value of q given p. Therefore $\Phi(p,q) = \{(p,q)|q \leq \phi(p)$ and $p+q \geq 1\}$. Of course $\frac{\partial\phi(p)}{\partial p}$ is negative, and we assume that the absolute value of the tradeoff between p and q is increasing in p, so that $p < p'$ if and only if $\frac{\partial\phi(p)}{\partial p} > \frac{\partial\phi(p')}{\partial p}$ (this is equivalent to $\frac{\partial^2\phi}{\partial p^2}$ being negative).

Figure 1 shows an example in which the technology set is symmetric around the reflection line $p = q$ (a condition we will continue to adopt).[8] For this case it is natural to define an information

[6] All proofs are presented in Appendix B.

[7] Because our empirical application entails looking at weather prediction over a relatively short time period, we do not consider changes in the set of feasible signals due to technological advancement.

[8] The symmetry assumption, however, is not necessary for the comparative statics we develop below.

structure's *bias* as $\mathbb{B} = p - q$. Here, when $p = q$, bias is zero, and the signal structure is designated *neutral.* The news provider can alternatively choose a signal structures for which $p > q$, in which case we have positive bias, which might also be called *optimistic* (since fair predictions now occur relatively more often than rainy predictions given equal realizations of weather). Signal structures where $q > p$ are *pessimistic.* Notice that in Figure 1, in line with the previous paragraphs, signals below $p + q = 1$ are not feasible (since they are equivalent to signals above that line), and signals above $\{(p, q)|q \leq \phi(p)\}$ are not available. In the figure, the example shown, (p^*, q^*), is a neutral case.

This set up corresponds in a natural way to our empirical work below. We look at bias $\mathbb{B}$ for a particular subset of days, say subset I, and then see if bias is different on an alternative subset J, i.e., we evaluate "differential bias" or "excess bias," in set I relative to J. $\Delta_{I,J} = \mathbb{B}_I - \mathbb{B}_J$. Evaluating excess bias is an important excercise when empirically testing the model. This is because there might be exogenous reasons why there might be bias — e.g., sunny weather might be easier to predict than rainy weather. However, these exogenous reasons should not vary across sets I and J. Thus, for example, we look to see if the *New York Times* weather reports exhibit different biases on home game days than on non-game days.

Intuitively, consumers who use weather predictions to condition decisions will prefer predictions that are more accurate; among structures in the feasible set, they will prefer structures on the boundary of the feasible set (as in the example shown in Figure 1). This idea — about the "informativeness" of the signal structure — is formalized in the following lemma:

Lemma 2 *Assume p and q jointly satisfy $p + q \geq 1$. Then (p', q') is Blackwell sufficient/more informative for (p, q) if and only if $p' \geq \frac{p}{1-q} - \frac{p}{1-q}q'$ and $p' \geq 1 - q'\frac{1-p}{q}$.*

If (p', q') is more Blackwell informative than (p, q) then the posteriors under (p', q') are a mean preserving spread of the posteriors under (p, q) — a result that follows from the law of iterated expectations. Clearly, this will be true if $p' > p$ and $q' > q$, but Lemma 2 shows it can also be true under less stringent conditions. Figure 2 illustrates using an example in which $(p, q) = (0.6, 0.6)$.

With all this in mind, we turn to the time line in our environment. In period 1 the decision-maker and news-provider share a prior ρ on the probability of a future state, in period 3, being A. A monopoly news provider picks an information structure and sets the price of the news. Consumers, knowing the information structure, and the price for information, choose whether to purchase information or not. In period 2, the news-provider receives a signal, and passes it along to their customers who purchases the information structure. Consumers choose an action. In period 3, the state is realized and consumers receive payoffs that depend on their actions and the realized state. Our restriction on the timing of actions is substantive, and we attempt to test this in data. We make this substantive restriction in order to provide the best case possible for a rational explanation of media bias.[9]

2.2.2 Preferences over Information Structures

Because no material payoffs occur before period 3, we model the decision-maker's environment as a compound (2-stage) lottery. Each set of payoffs $(H, L, 0)$, prior probability ρ, and an information structure (p, q) together generate a unique compound lottery. The technology constraint, along with the payoffs and prior, define a feasible set of compound lotteries. The information structure that leads to the highest expected utility for the decision-maker, given parameters, is denoted $(p^*(H, L, \rho), q^*(H, L, \rho))$. (We will typically suppress the dependence on the other parameter values to simplify notation.)

Expected utility maximizers receive no flow utility in periods 1 and 2, and so simply try to maximize the expected utility received in period 3. As noted above, if the agent takes action a, and the realized state is A, the payoff is H; if she takes action a and the realized state is B, the payoff is L; and if she takes action b, the payoff is 0 regardless of the realized state. Figure 3 demonstrates expected payoffs from each of the agent's action in terms of beliefs and utility. The horizontal axis represents the agent's belief in the probability of state A (fair weather). The vertical axis shows the expected payoff from actions conditional on those beliefs.

Using Figure 3 as an example, if the consumer has a prior $\rho = \frac{1}{2}$ and she receives no other information, her optimal action is a, i.e., she goes go to the game (since the average of L and H is greater than 0). (In the figure, her expected utility is designated "$u|$ no information"). Now suppose she receives weather predictions using signal structure S, given by $(p, q) = (1, \frac{1}{2})$. Using (1) and (2) it is easy to confirm that $\psi_R = 0$ and $\psi_F = \frac{2}{3}$, given a rainy or fair prediction, respectively. The consumer now conditions her action on the weather prediction, and this allows her to achieve an higher level of utility than in the absence of news.

Not all signal structures increase expected utility for the consumer. To see this, continue with the example from Figure 3, but consider the signal structure S' given by $(p, q) = (\frac{1}{2}, 1)$. Now posteriors

[9] Instead of using the assumption of "full commitment," where an agent has to choose an action in period 2, and so information is instrumentally valuable, we could make use of an alternative assumption that the agent makes her choice in period 3, after observing the realization of the state. In this latter case, information obviously has no instrumental value (i.e., it cannot help in decision making). Therefore neoclassical consumers have no value of information. Bias would result only from intrinsic preferences over beliefs and information. These two situations are obviously limiting cases of a more general model where the agent must make decisions in period 2 which are costly (either explicitly or implicitly) to realization in period 3.

generated are $\psi_R = \frac{1}{3}$ for a rainy signal and $\psi_F = 1$ for a fair signal. Here, regardless of which signal she observes, she will still go to the game. Therefore, her expected payoff from this information structure is the same as with no information. Clearly, then, our consumer prefers information structure S to S', since the former signal structure allows her to condition her actions and thereby increase the expected payoff. We now formalize these intuitions.

To explore circumstances under which a consumer will condition actions on signals, we observe, first of all, that expected utility for a consumer who conditions is

$$u = \rho p H + (1-\rho)(1-q)L. \tag{2.3}$$

Next notice that if the consumer does *not* condition, then her expected utility will be either (i) $\rho H + (1-\rho)L$ (if she always goes to the game), or it will be (ii) 0 (if she never goes to the game). So the consumer will want to condition behavior if expected utility (3) is greater than both object (i) and object (ii). This gives us two "conditioning constraints," only one of which is typically binding. Constraints (i) and (ii) can be characterized, respectively, as follows: in (p,q) space, signals must be above the lines,

$$q = -\frac{\rho H}{(1-\rho)L} + \frac{\rho H}{(1-\rho)L}p, \tag{2.4}$$

and

$$q = 1 + \frac{\rho H}{(1-\rho)L}p. \tag{2.5}$$

Notice that the first of the conditioning constraints passes through the point (1,0) while the second conditioning constraint passes through the point (0,1). Also observe that the constraints have the same slope (which is negative, as $L < 0 < H$), i.e., they are parallel. Thus, as noted, only one of the constraints is typically binding.

As long as there exists an information structure in the feasible set $\Phi(p,q)$ that meets the conditioning constraints, the consumer will have higher utility if she uses that information to condition her behavior. In this case, using (3), the agent has indifference curves defined over (p,q) pairs that satisfy

$$q = \frac{(1-\rho)L - \bar{u}}{(1-\rho)L} + \frac{\rho H}{(1-\rho)L}p \tag{2.6}$$

for various levels of $\bar{u}$. The slope of an indifference curve is negative. Indeed, the slope of an indifference curve is the same as the slope of each conditioning constraint.

Figure 4 illustrates. In this example, the point of tangency of the indifference curve and the technology boundary is an optimum if the conditioning constraints are met. It is easy to see that this point, (p^*, q^*), lies to the right of the outermost conditioning constraint (which is this example is the constraint that passes through (1,0)). Thus we do indeed have an optimal outcome for the consumer, in which she conditions her behavior on information.

Notice that the slope of the indifference curve becomes steeper as H, L, or ρ increase. The following proposition summarizes the resulting comparative statics.

Proposition 1 *If the decision-maker strictly prefers at least one signal structure to another, then p^* is increasing in H, L, and ρ, and q^* is decreasing in H, L, and ρ.*

2.2.3 Demand-Driven Bias

In order to examine information provision in equilibrium for our environment, suppose that on non-game days consumers have a preference for a neutral (i.e., $p = q$) information structure. On game days, consumers have the action set described in the previous subsections available to them. For the sake of simplicity we will consider a single representative consumer.

A monopolist firm can choose any feasible information structure on a given day (and the chosen information structure can vary by day), which it then sells at price r to consumers. The firm's profits are equal to r if it sells information and 0 if it does not (since the cost of of producing information is 0). Because the firm profits by selling the consumer her optimal signal, the firm will always provide the structure (p^*, q^*) discussed in the previous subsection.

The following proposition summarizes the fact that the information structure will generally be biased on game days (in the sense that only one particular set of parameter values generates a preference for unbiased signals). Furthermore, observed bias will change with the underlying parameters of the model in predictable ways: bias $\mathbb{B} = p^* - q^*$ is increasing in H, L, and ρ.

Proposition 2 *If any information is purchased on non-game days, the information structure produced will exhibit no bias. On game days, if any information is purchased, the information structure will generically be biased, and the bias will be increasing in H, L, and ρ.*

2.2.4 Supply-Driven Bias

In this section, we model bias as being driven by supply-side considerations, i.e., the firm has an incentive to alter consumers' beliefs. For example, a newspaper may be paid by a sports team to increase consumers' beliefs about the likelihood of fair weather. In this case, the newspaper would benefit the higher $\rho p + (1-q)(1-\rho)$ is. We consider a more general benefit function, $b(p, 1-q, L, H, \rho)$, where $b_i \geq 0$ for all $i \in \{p, 1-q, L, H, \rho\}$, $b_{ij} \leq 0$ for all $i, j \in \{p, 1-q, L, H, \rho\}$, and b is concave. The negative cross partials, and overall concavity capture two things: first, the fact that individuals with the lowest willingness to pay require the highest values of (p, q, L, H, ρ) to attend the game; and, second, that the baseball team (and so the firm) would only benefit to the extent that the stadium is not already selling out.[10] The firm also faces a cost of biasing information, $c(p-q)$, where c is convex in the amount of bias, and has a minimum at 0. This captures the fact that there are likely consumers who would prefer that the news not be biased.

The firm's marginal benefits are falling in the current amount of bias. Moreover, the marginal benefits to bias are also falling in ρ, H and L. Because the marginal costs are rising in the amount of bias, there will be less bias when ρ, H or L increase. This is true for any optimum of the firm.

Proposition 3 *When bias is supply-driven bias, if information is purchased on non-game days, the information structure produced will exhibit no bias. On game days, if any information is purchased, the information structure will always be positively biased and the bias will be decreasing in H, L, and ρ.*

2.2.5 Welfare

The impact of media bias on welfare can be ambiguous. Bias in and of itself does not harm consumers payoffs. If distortions are occurring for supply-side reasons, then consumers are worse off; they would prefer less bias. But, if distortions are generated on the demand side, then the bias is in fact optimal.

2.2.6 Testable Implications

The two models in this section have predictions not only about the direction of bias, but also about how the bias is affected by changes in the parameters. We will use these predictions to test the models in in the next section. The following table summarizes our theoretical results.[11]

[10] In the attendance records we have available, we find the stadium occasionally sells out.

[11] The insights of the particular models presented here are in fact more general. For example, any model that predicts that the information provider wants to encourage attendance at the game will generate comparative static predictions similar to our model of supply driven bias.

Model	Direction of Bias	Change in Bias as H Increases	Change in Bias as L Increases	Change in Bias as ρ Increases
Demand-Driven Bias	Either positive or negative	Increases	Increases	Increases
Supply-Driven Bias	Always positive	Decreases	Decreases	Decreases

In the empirical section we will also consider the predictions of other models of media bias, such as Gentzkow and Shapiro (2006) and Mullainathan and Shleifer (2005).

2.3 An Empirical Test: Weather Reporting by the *New York Times*, 1890–1899

In order to test predictions about media bias we use a novel data source — the daily weather reports given in the *New York Times* during the late nineteenth century. We test whether the *Times* was more likely to provide biased weather predictions on days when the New York Giants, the local professional baseball team, were scheduled to play a baseball game in the Manhattan Polo Grounds. To establish the plausibility of such bias, we begin with some observations about the historical context.

First, weather reporting, although improving, was still in a relative poor state. By the 1890s the practice of using almanacs and astrology to make weather predictions had practically disappeared, and the telegraph had made synoptic meteorology, or weather maps, a tractable method for making weather prediction. This new technique in weather forecasting heightened the usefulness and popularity of weather predictions in newspapers (Nebeker, 1995). Still, weather reporting was a highly subjective practice that relied on rough empirical rules, with a limited understanding of scientific facts and mathematical modeling. Nebeker indicates that even at the beginning of the 20th century, weather reporting was more of "an art rather than a science."

Second, during this period, the *New York Times* was a relatively small player in the the New York media market. There were over 10 newspapers competing for readers using morning, afternoon and evening editions, with prices for weekday editions typically less than three cents. The largest newspaper had a circulation of around 300,000 during the late 1800s.

During the period of analysis, 1890–1899, the *New York Times*, a morning newspaper, was in a transitional phase. In the spring of 1896, Adolph Ochs took over as owner (from Henry Raymond) of what was, at the time, a struggling and politically motivated newspaper. He instituted major changes, among them the change in weather reporting. He also coined the phrase "All the News That's Fit to Print." Ochs lowered the price of the paper from 3 cents to 1 cent, which perhaps was one reason for an increase in the circulation of the *Times*, which tripled within two years from 26,000 to 76,000.[12] Although we have some information about yearly average circulation, we do not have daily numbers, and so do not use circulation data in our analysis. The growth in circulation would also lead us to suspect a weakening in the motives to bias weather, as the circulation would grow beyond the local Manhattan market (where the Giants were located).[13]

As mentioned, under Ochs' new administration changes were made to weather reporting. Every morning the *Times* would provide the weather report for the rest of the day. Prior to June 16, 1896 the weather report quoted local weather predictions from the U.S. Weather Bureau. Although the Bureau was based in in Washington D.C., the Bureau had an office in Manhattan, and produced weather predictions for most major cities. After June 16, 1896 the *Times* switched its editorial policy and began producing in-house morning weather reports, as is indicated by the introduction of a separate column "Probabilities for the Day" on the front page of the paper. After the switch, forecasts no longer cited the Bureau as a source. We verified that the reports were in general not the same as other New York newspapers. The predictions by the *Times* during this period were simple. There was generally either a prediction of fair weather, or rain of some type (e.g., "showers" or "rain"). No probabilities were given, nor any anticipated amount of precipitation, nor typically any information about the timing of precipitation.

During the late nineteenth century, the *Times* had a sports section that provided extensive coverage of the National League and particularly the New York Giants, a major league baseball team that played in the Polo Grounds, in upper Manhattan. This was a period of rapidly growing popularity of baseball. New York had two major league teams, the Giants and the Brooklyn Bridegrooms (which later became the Dodgers), that were part of the National League, a 12 team league at the time. There were no other professional sports teams in the New York area at the time. The *Times* apparently catered to a Manhattan readership, as they provided extensive coverage of the Giants.

[12]The new weekday price of 1 cent is approximately 25 cents in today's terms. The Sunday edition was more expensive.

[13]In fact, bias does indeed empirically disappear. We collected data for 1910 and found no statistically significant evidence for bias in weather reporting associated with home game days.

The New York Giants were owned by businessmen Andrew Freeman and John Brush. Although we do not have data on price variation across years and games for tickets, we do know that a ticket price was 50 cents at the beginning of the time period we consider.[14] We cannot find any formal association between the *New York Times* and the Giants, nor association between their owners.

2.3.1 Data

To test the comparative statics and level of bias, we use a dataset constructed as follows: First, we coded home games during National League play for the Giants (typically late April to early October) for the 1890 through 1899 seasons.[15] The coding system is binary, with a 1 indicating a home game for the Giants and a 0 representing a non-home game (which includes away games). These records were taken from *Baseball Reference*, which has extensive coverage of historical records, including rankings and win-loss records.

Another part of the dataset is gathered weather reports from the *Times* during the period mentioned above. The *Times* weather report, indicating weather for the day, was published in the morning paper. We recorded an indicator variable — 1 for a fair prediction and 0 for a prediction which mentioned rain. As noted previously, weather reports were sufficiently simple so that there was a clear indication of how to code the data. Consider, as an example, the following weather forecast, given on the 17th of July, 1896:

Probabilities for To-day.
In this city: Fair, northwesterly winds.

Such a report is recorded in the data as a "fair" prediction. We were able to collect these data for all but four of the 1701 days in our study period.

The actual weather outcome was determined from records of the *National Climate Data Center of the National Center for Atmospheric Research*, which holds historical data from the Weather Bureau. We have daily precipitation levels at the Manhattan reporting station. The day was coded as "fair" if there was no rainfall (again, with 1 as an indicator), and "rainy" if precipitation was recorded (with 0 as the indicator).[16] We have data for all but two days.

For a relatively small number of days we were able to find attendance data from the *New York Times* sports page. The *Times* had a column the day after each home game summarizing the results,

[14] Approximately 10 dollars in today's terms.

[15] The first game of the season was between April 15 and 28, while the last day was between September 26 and October 15.

[16] The reporting station was at Central Park in Manhattan, very close to the Polo Grounds and also the the offices of the *New York Times.*

sometimes listing attendance. Attendance seems to be rounded to the nearest 500. We also collect the ranking of the New York Giants for each season, again using records from *Baseball Reference*.

2.3.2 Empirical Strategy and Results

Given the historical context, the goal is to test the models discussed in the previous section. After we discuss the results of our analysis, we relate them back to the predictions of the models. Doing so involves two steps. First, we ask whether the model can generate the direction of bias consistent with that observed in the data (e.g., positive bias). If so, we then evaluate comparative static predictions from the model. For instance, we look for correlations in the observed bias and variation in payoffs to attending a game or in priors in the likelihood of rain.

Before we turn to our primary analysis, we provide some summary statistics regarding two empirical objects of interest: *weather realizations* (as recorded by the Weather Bureau) and *weather predictions* (from the *New York Times*).

We begin with a description of weather realizations in Table 1. The top row gives statistics for the full sample (baseball seasons from 1890 to 1899), showing the proportion of days for which the realized weather was *fair* for each of the following subsets of days: first, all days; second, days when the Giants played a home game ("home game days"); third, days when there was not a home game ("non-home game days"); and, fourth, for completeness sake, for games played when the Giants were away ("away games"). As we have emphasized, our analysis focuses on the impact of policy changes in weather reporting at the *New York Times* on June 16, 1894. So, here we show how realized weather varied before and after that date.

As Table 1 shows, the weather is nicer on days with home games than on days that don't have home games. This is true in both the before period and the after period. There are two possible reasons for this. First, there were rainouts; these days, with no home game played, obviously would be exclusively rainy days. Second, it may be the case that the National League tended to schedule relatively fewer home games for the Giants in traditionally rainy months and relatively more games in fair months. In order to account for this fact, we will attempt to control for differences in priors across different days as a robustness check.

As for *weather predictions* from the *New York Times*, Table 2 provides summary statistics. Over the full sample (baseball seasons, 1890–1899), the proportion of *Times* weather reports predicting fair weather was 0.631, slightly lower than the corresponding realizations for fair weather realizations in Table 1. Given that the realized weather was fairer on home game days than on non-home game days, it is not surprising that predictions were on average fairer on home game days than on non-home game days. Below, we will be assessing bias by asking if, in the period after June 16, 1896, the *Times* tended to over-predict fair weather on days with a home game relative to days with no home game. Thus, it is interesting to note here that in the after period, fair weather predictions occurred with a substantially higher probability on home game days than on non-home game days.

We continue to summarize characteristics of our data by considering patterns of autocorrelation in weather outcomes and weather predictions. We do so in three regressions, reported in Table 3. First, Panel A examines possible autocorrelation between realized weather today and realized weather yesterday. This can be important as consumers might base their beliefs about fair weather on their observations about recent weather (i.e., beliefs might be formed using information beyond what the weather report predicts). We note that fair weather yesterday is associated with a significant increase (16 percentage points) in the probability of fair weather today. Panel B considers the same type of autocorrelation but looking at the *predicted weather* by the *New York Times.* Here we find a lower level of autocorrelation, just as we would expect if the weather predictions are a noisy guess about realized weather. Finally, Panel C provides information about the informativeness of the weather predictions above and beyond yesterday's weather realization. Put another way, if an individual observed yesterday's weather, would she still learn something from reading the *New York Times*? We find that both yesterday's weather and the weather prediction are informative about today's weather.

With these characteristics of the realized weather and predicted weather in mind, we now turn to our primary analyses, in which we evaluate the accuracy of predicted weather, conditional on the realized weather. Our goal here is to estimate the accuracy of *New York Times* weather predictions within the framework of the theory posited above. Thus we are interested in an empirical assessment of the probability of correctly predicting fair weather when in fact the weather will be fair (p), and the probability of correctly predicting rainy weather given the weather will be rainy (q).

Table 4 presents the observed values of p and q for various subsets of days. We note three features of the data. First, over the entire sample, the accuracy of predicting fair weather is somewhat higher than the accuracy of predicting rainy weather (0.754 compared to 0.624). This is true before the change to in-house weather predictions, on June 16, 1896, as well as after, as can be seen by comparing the second and fifth rows of the Table 4. Second, in the before period, when the Weather Bureau was producing weather reports, predicting accuracies p and q were quite similar on non-home game days and game days. Third, in contrast, in the after period p and q follow a pattern that is strikingly consistent with our model of optimistic bias on home game days relative to non-home game days. Thus p is higher on home game days than on non-home game days, and the converse is true for q. We note, in addition, that on non-home game days the *Times* used a somewhat different weather prediction policy (i.e., with lower p and higher q) than the Weather Bureau used in the before period.[17]

Figure 5 shows the information structures we observe: "before home" and "before non-home" (i.e., home game days and non-home game days before June 16, 1896) information structures look similar, while the corresponding "after home" and "after non-home" structures look very different. Consistent with the observations we have just made, this figure shows that although the before period weather reports for home game days were more Blackwell informative than the before period non-home game days reports, these differences were very slight. Moreover, no other rankings by Blackwell informativeness exist. It does not seem to be the case that the *New York Times* was receiving more Blackwell informative signals and then garbling them in two different ways on different days after the switch. Also, there appears to be a clear trade-off between p and q; high values of one accuracy are associated with lower values of the other accuracy. Finally, the figure shows that in the after period, weather reports on home game days are more optimistic than on non-home game days.

We now turn to a statistical test of our theory. Our primary question is: When the *New York Times* produced its own weather reports, was there "excess bias" on home game days relative to non-home game days, i.e., is $\Delta_{\text{After;Home,Non-Home}} = \mathbb{B}_{\text{Home}} - \mathbb{B}_{\text{Non-Home}} = (p_{\text{Home}} - q_{\text{Home}}) - (p_{\text{Non-Home}} - q_{\text{Non-Home}})$ positive? Using estimates from Table 4 we find

$$\Delta_{\text{After;Home,Non-Home}} = 0.291. \tag{2.7}$$

The standard error for this estimate of 0.087.[18] The t statistic is 3.48. We reject that $\Delta_{\text{After;Home,Non-Home}} = 0$ at the 0.001 level. Thus we have clear evidence of excess bias in the after period. In terms of our model, the game-day bias is optimistic; the *Times* tended to over-predict fair weather on home game days.

Next we try a counter-factual analysis. A potential concern with our finding of excess game-day bias in the after period is that it is being driven the fact that weather on home game days generally differed somewhat from weather on non-home game days (a fact documented in Table 1 for both

[17]We are not certain why the *Times* tended to adopt a more pessimistic weather reporting policy on non-home game days than the Weather Bureau. Our theory focuses only on how the *Times* might adopt differing prediction policies on home game days compared to non-home game days, and that also is the focus of our empirical work below.

[18]Given that outcomes shown in Table 4 are binomial, the standard error is calculated as follows: Sample sizes to estimate p_{Home}, q_{Home}, $p_{\text{Non-Home}}$, and $q_{\text{Non-Home}}$, respectively, are 185, 58, 233, and 155. So the standard error is

$$\sqrt{\frac{0.800(1-0.800)}{185} + \frac{0.500(1-0.500)}{58} + \frac{0.674(1-0.674)}{233} + \frac{0.665(1-0.665)}{155}} = 0.087.$$

the before and after period). This concern would be ameliorated if excess bias does not appear in the before period, when the *Times'* weather predictions came from the Weather Bureau.[19] In fact, we find that

$$\Delta_{\text{Before;Home,Non-Home}} = -0.002. \tag{2.8}$$

Excess home game bias is estimated to be very close to zero. This estimate is reasonably precise; the standard error is 0.065 (using the same approach as in footnote 17).

We can shed further light on our main results with regression analyses. Our baseline regressions, presented in the first column of Table 5 follows a differences-in-differences approach. Specifically, we estimate

$$\mathbb{D}_{\text{PredictCorrect}} = \beta_0 + \beta_1 \mathbb{D}_{\text{Fair}} + \beta_2 \mathbb{D}_{\text{Home}} + \beta_3 \mathbb{D}_{\text{Fair,Home}} + \epsilon. \tag{2.9}$$

In this regression, $\mathbb{D}$ are dummy variables. The dependent variable, $\mathbb{D}_{\text{PredictCorrect}}$, is 1 if the predicted weather was correct and 0 otherwise. $\mathbb{D}_{\text{Fair}}$ is 1 if the day was fair and 0 otherwise; $\mathbb{D}_{\text{Home}}$ is 1 if the Giants were playing at home, 0 otherwise; and $\mathbb{D}_{\text{Fair,Home}}$ is 1 for a fair day with a home game and 0 otherwise. The estimated coefficients in this regression are related to the p and q estimates given in Table 4: β_0 is an estimate of $q_{\text{Non-Home}}$, $\beta_0 + \beta_1$ is an estimate of $p_{\text{Non-Home}}$, and so forth. It can be shown that $\beta_3 = (p_{\text{Home}} - q_{\text{Home}}) - (p_{\text{Non-Home}} - q_{\text{Non-Home}})$; coefficient estimates of β_3 for the after and before periods correspond with estimates of excess bias given in (7) and (8) respectively.

Column (2) of Table 5 shows coefficient estimates of these same regressions also including year, month, and weekend fixed effects. The year effects allow for the possibility that accuracy varies over time. We have seen that the probability of fair weather varies by month, so we include month fixed effects in case this affects results. The inclusion of weekend indicator variables allows for the possibility that the bias was primarily for weekends and not home games more generally. Panel B shows that our key inference about excess bias in the after period is unchanged when we include these fixed effects; estimated excess bias (the coefficient on $\mathbb{D}_{\text{Fair,Home}}$) is about the same as in our baseline regression. For completeness sake, Panel A shows the result of this same regression in the before period — showing that inclusion of fixed effects does not alter our basic inference that there is no excess bias in the before period.[20]

[19] To be clear, we believe our theory potentially pertains when the *Times* makes in-house predictions, but make no claims one way or another about it's applicability in the before period.

[20] Finally, we estimated our baseline regression, reported in column (1) of Table 5, Panel B, but use a definition of "home" that includes only home games that were scheduled at the beginning of the season. (National League schedules were typically released in February or March prior to the season, and were published in the *New York Times* and other newspapers.) This definition thus includes home games that were rained out and excludes home games that were make-up games due to rain-outs. The basic inference is quite similar as with games actually played:

We next try an alternative approach, in which we specify a regression in which the dependent variable is the *Time's* weather *prediction* for a given day — 1 for fair and 0 for rain. We continue to use a differences-in-differences design:

$$\mathbb{D}_{\text{PredictFair}} = \alpha_0 + \alpha_1 \mathbb{D}_{\text{Fair}} + \alpha_2 \mathbb{D}_{\text{Home}} + \alpha_3 \mathbb{D}_{\text{Fair,Home}} + \epsilon. \tag{2.10}$$

Again, regression coefficients correspond to p and q for various subsets of days (as reported in Table 4), although the relationship is different than in our previous regression. Here α_0 estimates the probability that the *Times* predicts fair weather on a non-home day that is rainy, so $\alpha_0 = 1 - q_{\text{Non-Home}}$. Similarly, $\alpha_0 + \alpha_1 = p_{\text{Non-Home}}$, $\alpha_0 + \alpha_2 = 1 - q_{\text{Home}}$, and $\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 = p_{\text{Home}}$. It can be shown that excess bias is estimated by $2\alpha_2 + \alpha_3$. Column (1) of Panel B of Table 6 gives our baseline results; as expected we estimate $2\hat{\alpha_2} + \hat{\alpha_3} = 0.291$ (and as indicated by the F statistic reported in the note to Table 6, we reject the hypothesis that excess bias is 0).

Our reason for estimating regression (10) is that it allows us to ask what happens to our inference when we attempt to control for individuals' prior beliefs about the probability of fair weather on any given day. To do so, we first form estimates of individuals' prior beliefs by estimating a probit regression, in which the dependent variable is the weather realization, fair (1) or rainy (0) for a given day, and the independent variables are the amount of precipitation each of the previous five days. We then use results from this equation to generate the prior probability of fair weather, denoted *Daily Prior for Fair Weather*, using the information that would have been available to individuals before reading the weather report. One plausible theory of bias would be that there is an issue with prediction technology such that it forces p higher and q lower as the prior increases. This could account for the bias we observe. So we include this estimated prior as a control in regression (10).

Table 6 shows the coefficients from the probit specification used to construct priors. Column (1) is our basic specification. Column (2) adds week fixed effects for the baseball season, as a flexible way of picking up a possible persistent seasonal pattern in rain from late Spring through early Fall.[21] Finally, as a simple check, in column (3) we add a dummy variable for a game being scheduled that day, and discover that it is not statistically significant. We also notice that estimated coefficients on lagged daily rainfall are extremely similar across the specifications. The first of our specifications is the simplest, and it has the advantage that it is formed strictly with information that was available

The estimated coefficient on $\mathbb{D}_{\text{Fair,Home}}$ in the after period is 0.240 with a standard error of 0.079. In contrast, in the before period the estimated coefficient on $\mathbb{D}_{\text{Fair,Home}}$ is 0.005 with a standard error of 0.060.

[21] Specifically, we formed 26 dummy variables: for April 15-21, April 22-28, and so forth, up through the second week of October.

before individuals would be forming priors about a given day's weather.[22]

Estimates of regression (10) with and without the priors are reported in Table 7. Panel A gives results for the before period, simply as a counter-factual. Our key interest is Panel B, which gives analysis for the after period. We see that coefficients in our regression change very little across the two columns (other than the intercept) when we include the prior. When we include the prior in the regression, estimated excess bias is $2\hat{\alpha_2} + \hat{\alpha_3} = 0.248$ (and we reject the hypothesis that excess bias is 0). In short, it seems that our key inferences about *Times* weather reporting biases are not driven by differences in daily priors about the weather. As for Panel A—the counter-factual analysis for the before period—in both specifications we fail to reject the hypothesis that $2\alpha_2 + \alpha_3 = 0$.[23]

We can use our constructed prior to address two additional questions: First, whether weather predictions were informative beyond priors. Second, whether the game schedulers had information in excess of that of consumers regarding the weather, and if this could account for the fact that home game days were sunnier on average than non-home game days. (Alternatively, it may be the case that the National League tended to schedule relatively fewer home games for the Giants in traditionally rainy months and relatively more games in fair months.) Given our constructed daily priors for fair weather, we ran a regression estimating whether there is additional information provided by weather predictions, and the fact that a home game may be scheduled, in addition to the daily prior. We regressed the realized weather on the prior, the weather prediction, and dummy variables to indicate whether or not there was a home game scheduled. We find that the coefficients on the daily prior, 0.564 (s.e. = 0.164), and the weather prediction, 0.332 (s.e. = 0.028), are significant, indicating that both provide information regarding weather realizations. However, the coefficient on scheduled games is insignificant, indicating that knowing whether it was a home game day or not would not have further changed individuals' beliefs.

As a final robustness check, we undertake an additional counterfactual analysis, using weather reports from a different New York newspaper, the *Wall Street Journal*.[24] In the late nineteenth century, the *Journal* primarily covered financial topics, including American and international business and economic news; the paper's claim to fame was the construction and reporting of the Dow Jones "Average," one of the first index measures of stock prices on the New York Exchange. The

[22]In contrast, in the other two specifications, priors are formed using weather outcomes that have not yet occurred.

[23]Finally, we tried our analysis but with priors formed with the other two specifications given in Table 6. When we use the specification from column (2) to construct our prior we estimate bias to 0.226, and $F(1, 611) = 7.14$. When we use the specification from column (3) estimated bias is 0.190 and $F(1, 611) = 4.52$. In both specifications we reject the null hypothesis of no bias.

[24]This paper was established in 1882 by reporters Charles Dow, Edward Jones and Charles Bergstresser as a letter that provided financial news, and was converted by Dow Jones and Company into the *Wall Street Journal*, which distributed it's first issue in 1889.

paper did not typically cover sports, and so probably did not attract readers who wanted to learn news about the New York Giants. Like the *Times,* the *Wall Street Journal* provided a rudimentary weather forecast.

With this in mind, we coded the weather forecasts from the *Wall Street Journal* in the same fashion as for the *New York Times.* Our assumption is that the typical reader did not read the weather reports with the intention of using the information to attend games. Thus we would not expect that the *Journal's* reports to exhibit the same bias as the *Times.*

The *Journal* was an afternoon paper, which produced forecasts covering the next day; this was taken into account when analyzing predictions. We assembled data for the baseball seasons during the after period, 1896–1899. We collected *Journal* data for a total of 236 predictions; this is all that is available in digitized records, and certainly doesn't include the entire after period. For those 236 dates, one observation is missing in the *Times* records. We drop this day to focus only on dates when both papers made predictions, leaving a sample of 235 days.[25] For these dates we find that the *Times* produced excess bias of $\Delta = 0.365$ (which is not too different than for the entire sample), while the *Journal* had excess bias $\Delta = 0.017$, which is close to 0, as we would have expected.[26]

We next turn to some of the other predictions from our theory. Recall that our theory provides some reason to think that the bias will depend on the prior for fair weather (ρ) in the model.[27] We can look for such a relationship by looking at month-to-month variation in the prior. We start by using the realized weather in the sample to construct the probability of fair weather in any given month (as in Table 1). We denote these monthly priors $\hat{P}_{\text{Fair,Month}}$. (We drop April and October for the purpose of constructing these priors as those months have few observations.) Figure 6 shows that for the post-June 1896 period, there is a clear negative relationship between these priors and the average excess game day bias in that month, $\Delta_{\text{AfterHome,AfterNon}-\text{Home,Month}}$.[28] In Figure 6 the

[25]The two papers agree on 166 observations (70.6% of the 235).

[26]These estimates of the *Times'* excess bias and *Journal's* excess bias are not statistically different at the 0.05 level. (The sample size is small, so the test is under-powered.)

[27]If individuals' beliefs are rational, we would expect a one percent increase in beliefs about the probability of fair weather to lead to a one percent increase in the probability of fair weather being realized. To be clear, these beliefs would be individuals' priors and posteriors, not the weather predictions themselves (although posteriors would be constructed using priors and the weather predictions). We conduct such a test, first regressing realized weather on our constructed daily priors, and second regressing realized weather on a constructed measure of a daily posterior, which we create by applying Bayes Rule, using the prior and weather predictions, allowing for different values of p and q across different subsets of days (the interaction of before, after, home and not home). Both regression coefficients are tightly estimated to be 1, in line with the rational expectations literature.

[28]One might be concerned that individuals' priors for game days and for non-game days within a month differed. Instead of lumping game days and non-game days within a month together, one could instead compute priors for game days and non-game days for each month separately, and then estimate the effects of priors on excess bias. Such an analysis again indicates that excess bias is falling (at a significant level) in the prior.

relationship appears to be roughly linear; we thus fit a line to these observations,

$$\Delta_{\text{AfterHome,AfterNon-Home,Month}} = \beta_0 + \beta_1 \hat{P}_{\text{Fair,Month}} + \epsilon,$$

where $\Delta_{\text{AfterHome,AfterNon-Home,Month}}$ is, for a given month, the difference in bias between home game days and non-home game days. Note that this regression is not simply picking up effects of summer relative to non-summer months, nor months early in the season relative to late in season (since fair weather is not perfectly correlated with either of these). We estimate β_1 to be -3.68 (with a standard error of 1.02), which is a statistically significant relationship at the 0.05 level.

Next recall that our theory suggests that bias will be related to the value fans place on being at a game when the weather is fair (H) and on loss from attending a game when it is rainy (L). Intuitively, these values might be related to the relative success of the team. As seen in Figure 7, in fact relative bias was highest in 1897, when the Giants had a relatively good season; was lower in 1898 and 1896, when the Giants finished 5th (of the 12 teams in the league); and was lower yet in 1899 when the team finished near the bottom of the standings.[29]

2.3.3 Further Analysis: Attendance

In order to get a better sense of how predicted and realized weather might affect the actions of readers and fans, we analyze the effects of weather predictions and realizations on the New York Giant's home game attendance. It also seems plausible that the performance of the team should affect attendance of the game when the team is doing well, and so we also look at the impact of team ranking on attendance. We estimate the following regression for the period June 16, 1896 and after:

$$A = \beta_0 + \beta_1 \mathbb{D}_{\text{PredictFair}} + \beta_2 \mathbb{D}_{\text{Fair}} + \sum_i \beta_i \mathbb{D}_{\text{Rank}} + \epsilon,$$

where A is the game attendance, $\mathbb{D}_{Rank}$ are dummy variables for the ranking of the team (either rank 3, 5, or 10 depending on the year) and the rest of the variables are as described previously. Recall that we only have a relatively small subset of games with attendance records, and as such take these results as preliminary.[30] Table 8 reports the results of this regression (with rank 5 being the

[29]The difference in estimated excess bias for the year with the team finished 3rd (1897) and year it finished 10th (1899) is statistically significant at the 0.05 level. (No other differences between years were statistically significant.) Of course the year-to-year differences in excess bias may have had little to do with the ranking. Our only point here is that this evidence is not what we would have predicted with our model.

[30]42.0% of the attendance records are complete. The *Times* provided a report the day after every home game, only some of which included attendance records. Efforts are being made to collect more data.

omitted category). Surprisingly, predicted weather is not a significant predictor of attendance. Even more surprisingly, realized weather is not correlated with attendance either. The only significant variable is the yearly rank of the New York Giants.

2.3.4 Relating Evidence and Theory

In the previous section we derived some testable implications of two models of media bias. We revisit them here:[31]

Model	Direction of Bias	Change in Bias as H Increases	Change in Bias as L Increases	Change in Bias as ρ Increases
Demand-Driven Bias	Either positive or negative	Increases	Increases	Increases
Supply-Driven Bias	Always positive	Decreases	Decreases	Decreases

We believe that it is reasonable to use the monthly average probability of a rainy day as a proxy for consumer's priors about fair weather in a given month. Furthermore, we believe that if the New York Giants were doing better, it represented an increase to H (i.e., it is better to watch a good baseball team rather than a bad baseball team when the weather is sunny).

Given these interpretations, recall that we found that as the prior increases the bias shrinks. This is consistent with supply-driven bias, but not with demand-driven bias. In contrast, we also found that as the team did better, the bias grew, which is inconsistent with supply-driven bias, but is consistent with demand-driven bias. In fact, neither of the models correctly predict both observed comparative statics.

Casting further doubt on both the demand- and supply-driven stories is our evidence that the predicted weather does not influence attendance. For example, if the team was paying the newspaper to bias the weather reports in hopes of boosting attendance, they would have presumably quickly

[31] One interesting alternative model, which we thank David Gill for mentioning, is that on game days the information is meant to better predict the weather only at the baseball game. On a non-game day, rain would be the prediction if it rained at all, but on a game day, rain would only be the prediction if it rained during the game. This could generate positive bias as we observe in the data. However, if days were particularly rainy, then the bias should shrink relative to non-game days, as it is more likely that on the rainiest days, it is also raining during the game. We observe the opposite relationship in the data.

learned that this was not effective.

Of course, there are also other existing models of media bias. Gentzkow and Shapiro (2006) develop a model where firms bias news for reputational reasons, i.e., in order to appear informed. Their intuition would predict that in our setting, as consumer's priors increase, the bias should grow. In fact, we observe the opposite. Mullainathan and Shleifer (2005) have a model of media bias where bias is driven by consumers' desires to see their beliefs confirmed. Although their model is not easy to put into our setting, their results suggest again that as consumer's priors increase, the bias should grow, which is opposite to what we observe.

Of course, we have made two assumptions in deriving the results of the models. First, individuals will condition their actions on the predictions if optimal. Second, individuals know (at some level) the values of p and q on non-home game days and home game days, and change their updating rules accordingly. Violating the first assumption would eliminate the prediction of bias in the two models we derive in this paper, not changing our overall conclusions. Violating the second assumption would immediately imply that there would be no benefit for the *New York Times* to biasing their information in order to improve consumers' ability to condition their actions, and so we would not expect to see bias on game days. Furthermore, violating the second assumption would not change the comparative static results in the case of supply-driven bias. Therefore, our conclusions would also not be changed by violating the second assumption.

2.4 Conclusion

Our data, consisting of local weather predictions from the *New York Times* and realized weather in Manhattan during the years 1890-1899, provide a new setting to try to understand media bias. We find evidence that the *New York Times* biased its weather reports on home baseball game days relative to non-home baseball game days when it produced its own weather reports. We find that the excess bias on home game days may fall when the ex-ante probability of sunny weather grows. We also find that the excess bias on home game days may be higher when the baseball team does better. These comparative statics are inconsistent with the rational models of media bias that we set out in our paper. Moreover, in the limited subset of data for which we have attendance data, it does not appear that predicted weather influences attendance, which would indicate that individuals do not condition their actions on the information provided by the weather prediction.

The data has several advantageous features. First of all, because we know both the weather prediction and weather realization, we can construct a measure of absolute (rather than comparative) bias. Second, the realized weather is unaffected by the weather report, and so there are no issues with endogeneity. Third, our data allow us to measure how individuals' posteriors should change as the bias grows or falls.

Of course, other mechanisms can also generate media bias, including models where consumers have preferences over beliefs or changes in beliefs. Examples include:

- Brunnermeier and Parker (2005) develop a model in which agents can choose their priors. They receive current flow utility from having high beliefs (which are based off their priors) about future flow payoffs. However, agents may take sub-optimal actions based on incorrect beliefs (since their beliefs are based on their chosen priors, not on true priors). Ex-ante, decision makers choose priors to maximize the weighted sum of the flow consumption utility and flow belief-based utility. However, if information has no instrumental value, then Brunnermeier and Parker would predict no need for biased information.

- Koszegi and Rabin (2009) derive a model where decision-makers receive both flow utility from consumption and also gain-loss utility that results from changes in beliefs about current and future flow consumption utility. Furthermore, in order to capture loss-aversion, individuals experience losses more strongly than they feel gains.

- In a similar vein, Artstein-Avidan and Dillenberger (2010) extend Gul's (1991) model of disappointment aversion to a dynamic setting.

- Kreps-Porteus (1978) preferences, and the parametrization considered by Epstein and Zin (1989) capture consumers who have a preference for early versus late resolution of information.

However, not all models which incorporate beliefs can capture the observed data. We have found some parameterizations of Koszegi and Rabin (2009) that can accommodate the data, and we are currently engaged in ongoing work to understand the generality of such results. However, the predictions generated by any of these preferences can change depending on several key assumptions.

For example, in models in this paper we assumed that information held instrumental value for consumers, i.e., consumers want to condition their actions on the weather report they observe. Testing the validity of this assumption is difficult. But, importantly, in the subset of days for which we had attendance data, we found that that weather predictions did not affect attendance. This lends support for the idea that consumers likely have non-expected utility preferences. Of course, whether or not consumers do condition their actions on the signals also has implications for the predictions of models in which consumers have non-expected utility preferences.

We also assumed that individuals know the values of p and q on non-home game days and home game days, and so will have different posteriors (given the same starting prior) from a fair weather signal depending on what kind of day it is. One could instead assume that the consumers think that all days are governed by the same p and q. Unfortunately, testing this assumption is very difficult, and there is nothing in our research design that could shed light on the answer.

Although we attempt to test the predictions of some models of media bias, we caution that it is not clear how far the intuitions and results in this paper extend. Our environment features information about outcomes that are easily verifiable within a short span of time. In many situations outcomes will not be verified by individuals receiving information (for example, information about climate change). There are many other situations where the information is about causal claims that involve a stochastic data generating process (for example, that smoking causes lung cancer). In both cases, it is easier for individuals to hold distorted beliefs. In these situations other motivations for media bias may dominate.

We would like to conclude by reflecting on the welfare implications of media bias. If individuals were conditioning their actions on the weather prediction, bias on game days could be welfare improving for individuals who may want to attend the game. In contrast, if media bias was being driven by the *New York Times* wanting to increase attendance, then bias could reduce welfare for fans. Of course, we do not find support for either of these models. In contrast, in those models where individuals receive utility directly from beliefs or changes in beliefs, such as those which can potentially rationalize our data, bias can be welfare improving (or harming) even if individuals do not condition their actions on the signal realizations. Furthermore, bias could be welfare improving even if it causes individuals to take actions that lead to worse material payoffs.

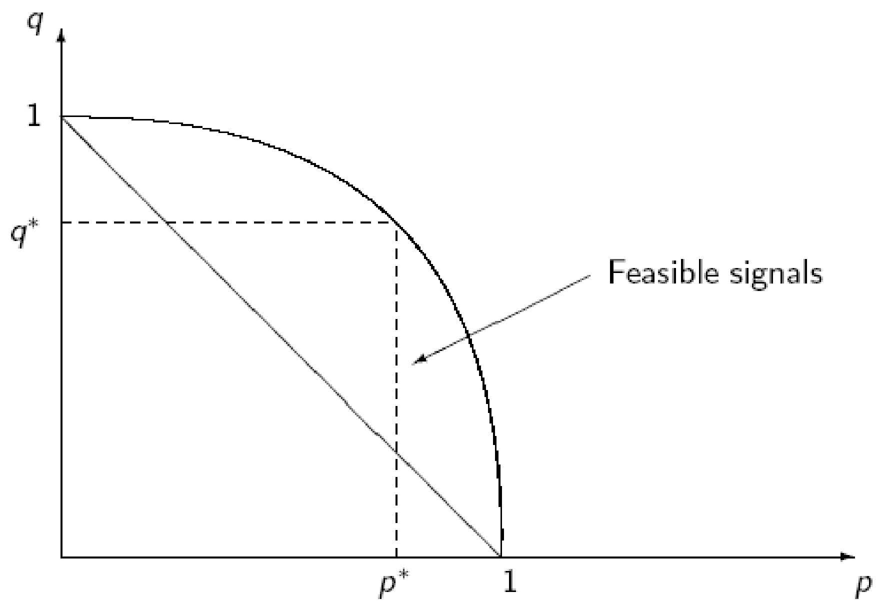


Figure 2.1: The Feasible Set of Information Structures

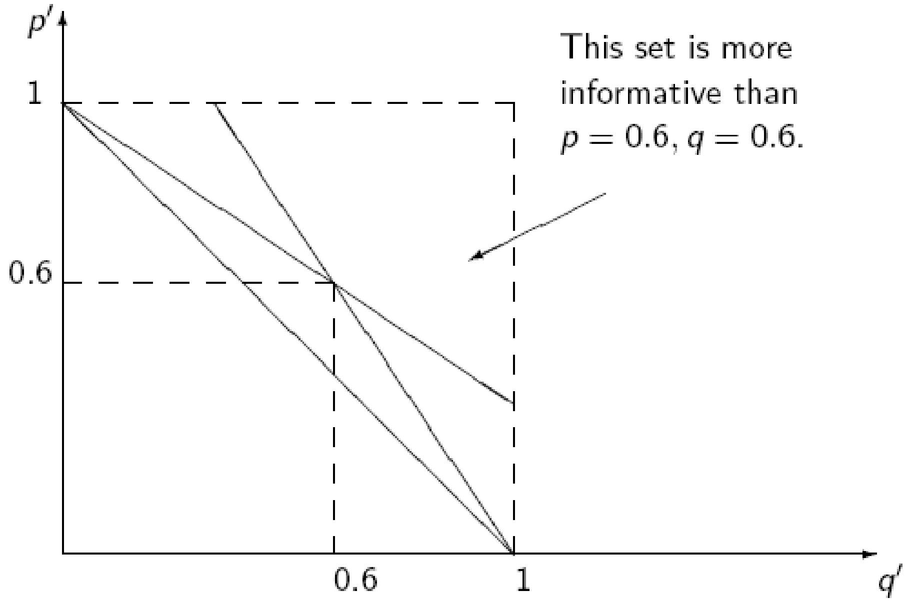


Figure 2.2: Signals that are Blackwell More Informative than (0.6,0.6)

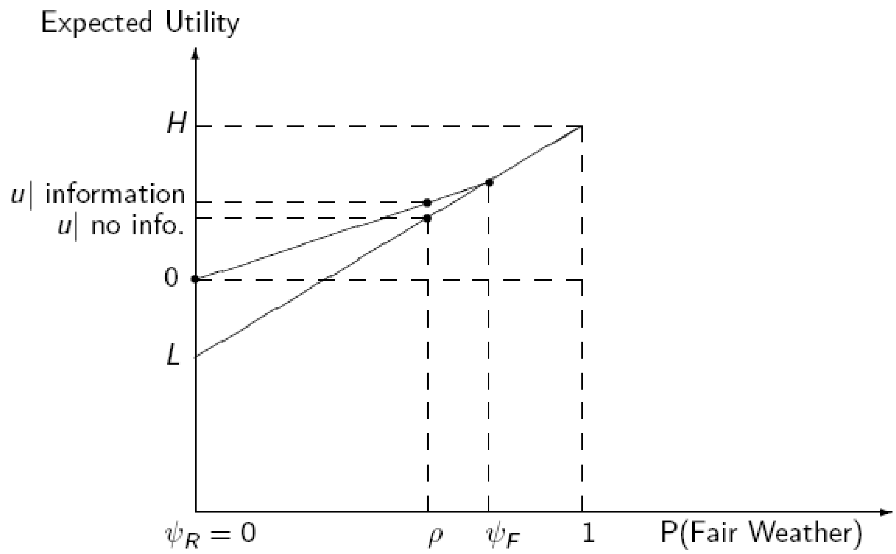


Figure 2.3: Expected Utility Payoffs with and without Information

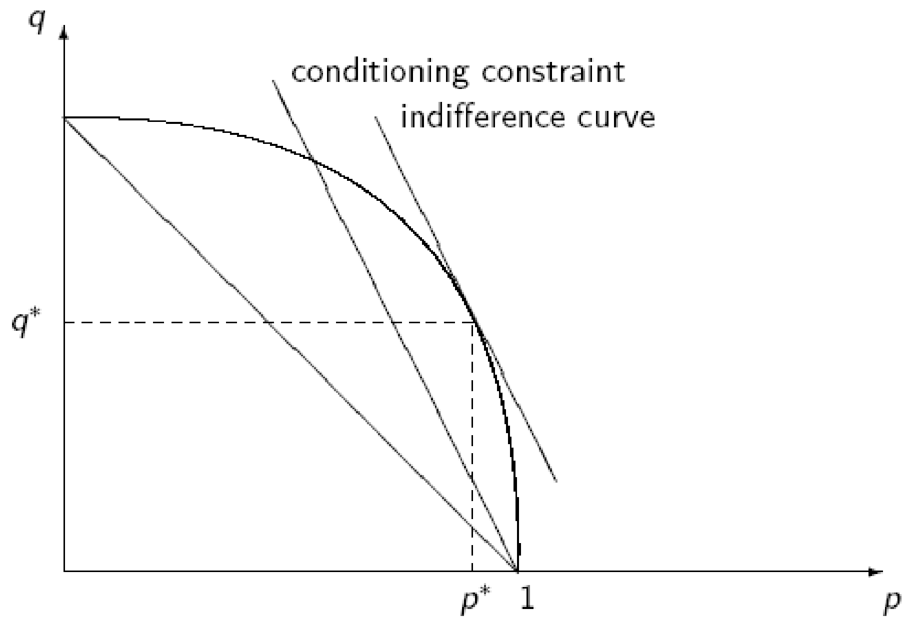


Figure 2.4: An Example of a Utility-Maximizing News Structure

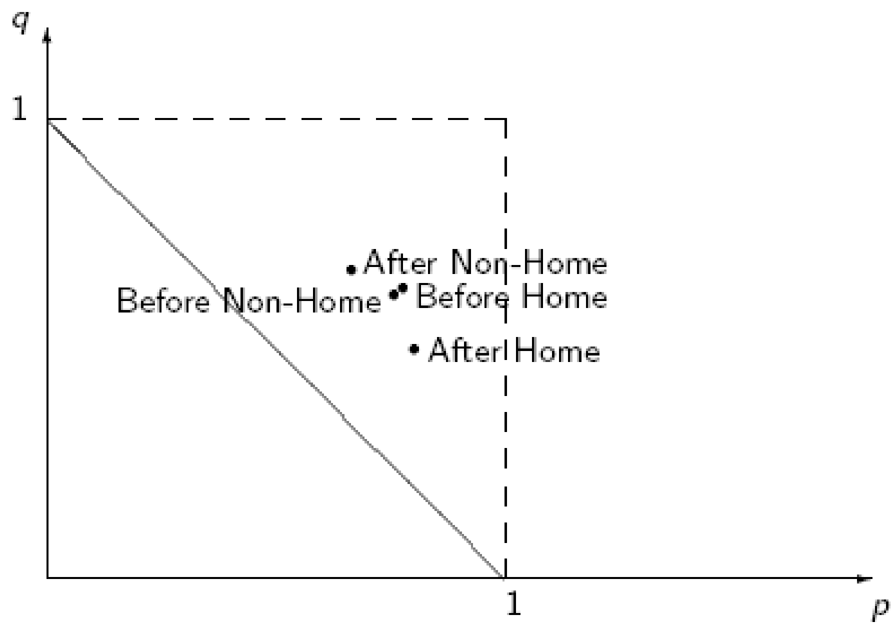


Figure 2.5: Empirically Observed Information Structures

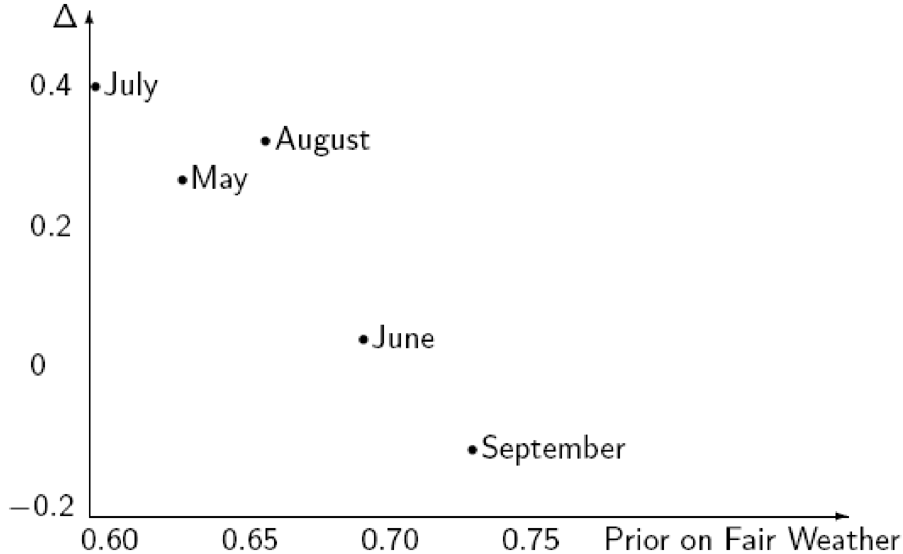


Figure 2.6: Relationship Between Monthly Priors for Fair Weather and Excess Bias Δ

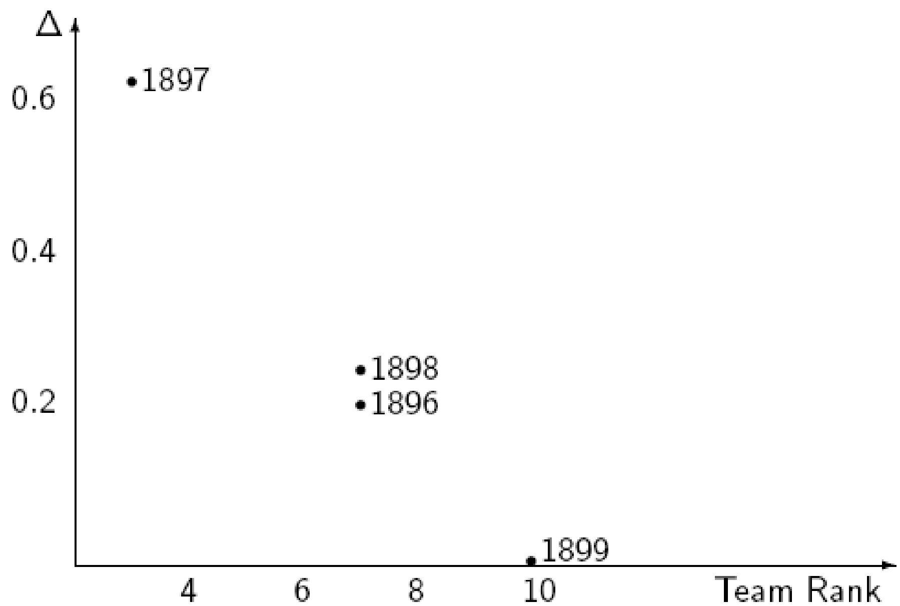


Figure 2.7: Relationship Between Team Ranking and Excess Bias Δ

Table 2.1: Summary Statistics: *Realized Fair Weather*

	All Days	Days with Home Game	Days with No Home Game	Days with Away Game
Full Sample, 1890 to 1899	0.676	0.755	0.626	0.678
Number of Days (n)	(1695)	(650)	(1045)	(642)
Dates before June 16, 1896	0.683	0.752	0.641	0.700
(n)	(1064)	(407)	(657)	(406)
June 16, 1896 and After	0.662	0.761	0.601	0.640
(n)	(631)	(243)	(388)	(236)
By Month:				
April	0.732	0.775	0.711	0.864
(n)	(123)	(40)	(83)	(44)
May	0.612	0.678	0.586	0.608
(n)	(309)	(87)	(222)	(143)
June	0.692	0.772	0.603	0.667
(n)	(299)	(158)	(141)	(81)
July	0.643	0.756	0.612	0.627
(n)	(308)	(82)	(226)	(150)
August	0.650	0.692	0.614	0.673
(n)	(309)	(143)	(166)	(104)
September	0.752	0.845	0.684	0.763
(n)	(294)	(123)	(171)	(97)
October	0.736	0.824	0.694	0.783
(n)	(53)	(17)	(36)	(23)

Note: Authors' calculations, data collected from the *National Climate Data Center of the National Center for Atmospheric Research* for the baseball season, 1890–1899.

Table 2.2: Summary Statistics: *Predicted Fair Weather*

	All Days	Days with Home Game	Days with No Home Game	Days with Away Game
Full Sample	0.631	0.695	0.591	0.615
(n)	(1695)	(650)	(1045)	(642)
Before June 16, 1896	0.643	0.676	0.623	0.643
(n)	(1064)	(407)	(657)	(406)
June 16, 1896 and After	0.612	0.728	0.539	0.568
(n)	(631)	(243)	(388)	(236)

Note: Authors' calculations, data collected from the *New York Times* for the baseball season, 1890–1899. **significance = 0.01. *significance = 0.05.

Table 2.3: Regression Results: *Relationships Between Realized Weather, Past Weather, and Predicted Weather*

A. Dependent Variable: *Realized Fair Weather* (1 if Fair, 0 Else)	
Constant	0.564**
	(0.020)
Realized Fair Weather Yesterday	0.164**
	(0.024)
n	1685

B. Dependent Variable: *Predicted Fair Weather* (1 if Fair, 0 Else)	
Constant	0.555**
	(0.019)
Predicted Fair Weather Yesterday	0.119**
	(0.024)
n	1685

C. Dependent Variable: *Realized Fair Weather* (1 if Fair, 0 Else)	
Constant	0.385**
	(0.026)
Realized Fair Weather Yesterday	0.116**
	(0.035)
Predicted Fair Weather	0.327**
× Realized Fair Weather Yesterday	(0.028)
Predicted Fair Weather	0.358**
× Realized Rainy Weather Yesterday	(0.037)

Table 2.4: Summary Statistics: *Times Accuracy of Weather Reports*

	p	q
Total	0.754	0.624
	(1145)	(550)
Before June 16, 1896		
All Days	0.768	0.626
	(727)	(337)
Non-Home Games	0.758	0.619
	(421)	(236)
Home Games	0.781	0.644
	(306)	(101)
June 16, 1896 and After		
All Days	0.730	0.620
	(418)	(213)
Non-Home Games	0.674	0.665
	(233)	(155)
Home Games	0.800	0.500
	(185)	(58)

Note: Authors' calculations, data collected from Baseball Reference, *New York Times*, and the *National Climate Data Center of the National Center for Atmospheric Research* for the baseball season, 1890–1899.

Table 2.5: Regression Results: *Times Correct Weather Prediction*

A. Before June 16, 1896	(1) No Fixed Effects	(2) With Fixed Effects
Constant	0.619**	—
	(0.029)	
Fair	0.139**	0.142**
	(0.036)	(0.036)
Home	0.025	0.034
	(0.053)	(0.054)
Fair × Home	-0.002	-0.006
	(0.063)	(0.063)
Year, Month, Weekend F.E.?	no	yes
n	1064	1064
R^2	0.022	0.038

B. June 16, 1896 and After	(1) No Fixed Effects	(2) With Fixed Effects
Constant	0.664**	—
	(0.037)	
Fair	0.009	0.006
	(0.047)	(0.047)
Home	-0.165*	-0.200**
	(0.070)	(0.071)
Fair × Home	0.291**	0.294**
	(0.083)	(0.083)
Year, Month, Weekend F.E.?	no	yes
n	631	631

Table 2.6: Probit Regression Results: *Construction of Daily Priors for Fair Weather*

	(1) Without Week Fixed Effects	(2) With Week Fixed Effects	(3) With Week Fixed Effects
Constant	0.468	–	–
	(0.041)		
Rain in Day $t-1$	-0.443**	-0.424**	-0.420**
	(0.097)	(0.099)	(0.099)
Rain in Day $t-2$	0.206	0.218*	0.222*
	(0.109)	(0.111)	(0.111)
Rain in Day $t-3$	0.008	0.014	0.020
	(0.102)	(0.104)	(0.104)
Rain in Day $t-4$	0.140	0.140	0.147
	(0.107)	(0.110)	(0.110)
Rain in Day $t-5$	-0.011	-0.027	-0.025
	(0.099)	(0.101)	(0.101)
Home Game Scheduled	–	–	0.131
			(0.072)
n	1635	1635	1635
Pseudo R^2	0.012	0.030	0.031

Note: Author's calculations, data collected from *New York Times* (for the scheduled home games), and the *National Climate Data Center of the National Center for Atmospheric Research* for the baseball season, 1890–1899. 26 week fixed effects are included in regression (2). **significance = 0.01. *significance = 0.05.

Table 2.7: Regression Results: *Times Predict Fair Weather*

A. Before June 16, 1896	(1) Without Prior	(2) With Prior
Constant	-0.025**	-0.440**
	(0.053)	(0.167)
Fair	0.376**	0.353**
	(0.036)	(0.037)
Home	$-0.025^{\dagger}$	$-0.040^{\ddagger}$
	(0.053)	(0.053)
Fair × Home	$0.048^{\dagger}$	$0.070^{\ddagger}$
	(0.062)	(0.063)
Daily Prior for Fair Weather	–	1.241**
		(0.250)
n	1064	1019
R^2	0.129	0.150

B. June 16, 1896 and After	(1) Without Prior	(2) With Prior
Constant	0.335**	-0.277
	(0.037)	(0.192)
Fair	0.338**	0.320**
	(0.047)	(0.048)
Home	$0.165^{*\dagger\dagger}$	$0.121^{\ddagger\ddagger}$
	(0.070)	(0.071)
Fair × Home	$-0.038^{\dagger\dagger}$	$0.006^{\ddagger\ddagger}$
	(0.083)	(0.083)
Daily Prior for Fair Weather	–	0.933**
		(0.285)
n	631	616

Table 2.8: Regression Results: *Attendance, for the Period after June 16, 1896*

Constant	4711.43**
	(1239.50)
Predict Fair Weather	341.68
	(1068.31)
Realized Fair Weather	-463.36
	(1026.86)
Year when the Rank was 3	2289.51*
	(1003.01)
Year when the Rank was 10	-1015.84
	(1238.75)
n	102
R^2	0.102

Note: Authors' calculations, data collected from Baseball Reference, *New York Times*, and the *National Climate Data Center of the National Center for Atmospheric Research* for the baseball season from June 16, 1896 through 1899. **significance = 0.01. *significance = 0.05.

Lemma 1 *The set* $\{(p,q)|(p,q) \in [0,1] \times [0,1]$ *and* $p+q \geq 1\}$ *satisfies three properties:*

1. *Observing a fair signal increases the posterior on state* A *relative to the prior, and observing a rainy signal decreases the posterior on state* A *relative to the prior.*

2. *For any signal structure* $(p',q') \in [0,1] \times [0,1]$*, there exists a* $(p,q) \in \{(p,q)|(p,q) \in [0,1] \times [0,1]$ *and* $p+q \geq 1\}$ *that generates the same posteriors with the same probabilities as* (p',q')*.*

3. *For any strict subset* S *of* $\{(p,q)|(p,q) \in [0,1] \times [0,1]$ *and* $p+q \geq 1\}$*, there exists a point* $(p',q') \in [0,1] \times [0,1]$ *such that there is no element of* S *that generates the same posteriors as* (p',q')*.*

Proof We will prove each part of the Lemma in turn. First we prove the first part. Recall that for a given prior $0 < \rho < 1$ on state A (fair weather) and information structure (p,q), the posterior for A given the fair signal is

$$\psi_F = \frac{\rho p}{\rho p + (1-\rho)(1-q)}.$$

Now $\psi_F > \rho$ if and only if

$$\psi_F = \frac{\rho p}{\rho p + (1-\rho)(1-q)} > \rho,$$

which holds if and only if

$$(1-\rho)p > (1-\rho) - (1-\rho)q,$$

which is the same as

$$p + q > 1.$$

An analogous series of steps establishes the result for the posterior after observing a rainy signal,

$$\psi_R = \frac{\rho(1-p)}{\rho(1-p) + (1-\rho)q}.$$

To prove the second part, assume that $p + q \leq 1$. In this case, denote $p' = 1-p$ and $q' = 1-q$. We will simply work with likelihood ratios. Under (p,q), likelihood ratio $\frac{p}{1-q}$ occurs with probability $\rho p + (1-\rho)(1-q)$ and likelihood ratio $\frac{1-p}{q}$ occurs with probability $\rho(1-p) + (1-\rho)q$.

Under (p',q') likelihood ratio $\frac{1-p'}{q'} = \frac{p}{1-q}$ occurs with probability $\rho(1-p') + (1-\rho)q' = \rho p + (1-\rho)(1-q)$. Likelihood ratio $\frac{p'}{1-q'} = \frac{1-p}{q}$ occurs with probability $\rho p' + (1-\rho)(1-q') = \rho(1-p) + (1-\rho)q$. Therefore (p',q') generates the same posterior distribution as (p,q). Moreover, $p' + q' = (1-p) + (1-q) = 2 - p - q \geq 1$ since $p + q \leq 1$. So therefore, instead of considering some (p,q) we can always instead consider the corresponding $p' = 1-p, q' = 1-q$. This proves the second part.

To prove the third part, observe that in order for two signal structures (p, q) and (p', q') to generate the same posteriors (so that for both signal structures, a fair weather prediction increases the posterior relative to the prior *and* a rainy weather prediction decreases it) it must be the case that $\frac{p'}{1-q'} = \frac{p}{1-q}$ and $\frac{1-p'}{q'} = \frac{1-p}{q}$.

Therefore $p' - p'q = p - pq'$ and $q - p'q = q' - pq'$, which is equivalent to $q = \frac{-p+pq'+p'}{p'}$ and $q = \frac{q'-pq'}{1-p'}$. Simplifying, we have $\frac{-p+pq'+p'}{p'} = \frac{q'-pq'}{1-p'}$, or $p'q' - pq'p' = -p + pq' + p' + pp' - pp'q' - p'^2$. This implies that $p'q' = -p + pq' + p' + pp' - p'^2$, and so $p(1-q'-p') = -p'q' + p' - p'^2 = p'(1-q'-p')$, which holds only if and only if $p = p'$. This proves the third part.

□

Lemma 2 *Assume p and q jointly satisfy $q \geq 1-p$. Then (p', q') is Blackwell sufficient/more informative for (p, q) if and only if $p' \geq \frac{p}{1-q} - \frac{p}{1-q}q'$ and $p' \geq 1 - q'\frac{1-p}{q}$.*

Proof Recall that one signal structure (p', q') is Blackwell more informative than another (p, q) if and only if the distribution of posteriors induced by (p', q') is a mean preserving spread of the distribution induced by (p, q). By the law of iterated expectations, the expected posterior under (p', q') and (p, q) must be the same — the prior. Because there are only 2 signals, there will be only 2 posteriors. So we just have to show that the posteriors under (p', q') are more extreme (in the sense that they are farther from the prior) than the posteriors under (p, q). In order to simplify the proofs, we will show an equivalent result — that the likelihood ratios under (p', q') are more extreme (farther from 1) than the likelihood ratios under (p, q).

The likelihood ratios after observing a fair signal under (p', q') and (p, q) are (respectively) $\frac{p'}{1-q'}$ and $\frac{p}{1-q}$ while the likelihood ratios after observing a rainy signal are $\frac{1-p'}{q'}$ and $\frac{1-p}{q}$.

In order for the ratios under (p', q') to be farther from 1 than (p, q), then $\frac{p'}{1-q'} \geq \frac{p}{1-q}$ and $\frac{1-p'}{q'} \leq \frac{1-p}{q}$. This is equivalent to $p' \geq \frac{p}{1-q} - \frac{p}{1-q}q'$ and $p' \geq 1 - q'\frac{1-p}{q}$.

□

Proposition 1 *If the decision-maker strictly prefers at least one signal structure to another, then p^* is increasing in H, L, and ρ and q^* is decreasing in H, L, and ρ.*

Proof Because any feasible signal structure is Blackwell dominated by a signal structure on the technology constraint, the agent will always want to choose an information structure along the technology constraint, if any signal feasible signal structure allows her to condition her action. If the agent finds it optimal to condition, then she must be getting a higher payoff than by not conditioning.

First we deal with cases in which the consumer conditions her actions. This occurs when neither conditioning constraint, (4) or (5), is binding. Recall that indifference curves have slope $\frac{\rho H}{(1-\rho)L}$.

This slope is negative (because $L < 0$), and is decreasing in H, L, and ρ. If the agent conditions her actions, and chooses an interior solution (p^*, q^*) at the point of tangency to the boundary of the technologically-feasible set (as in Figure 4), then standard comparative statics lead to the result that p^* is increasing in H, L, and ρ, while q^* is decreasing in H, L and ρ.

Next, we consider cases that lead the consumer to *not* condition her actions. As discussed in the text, this happens when one of the two conditioning constraints, (4) or (5), holds with equality (i.e., is binding). Notice that the slope of the conditioning constraints is the same as the slope of the indifference curves for a consumer who conditions her actions, and recall that the binding conditioning constraint passes either through $(p, q) = (1, 0)$ or $(p, q) = (0, 1)$. So if the consumer is not conditioning her actions, she must be at one of these two corners; either the consumer's indifference curve is so steep that her optimal signal structure, among feasible signals, is $(1, 0)$, or so flat that her optimal signal structure is $(0, 1)$. Either way, the conditioning constraint holds with equality, so the consumer places zero value on any information structure available in the feasible set. Such a consumer is thus trivially indifferent over signal structures.

$\square$

Proposition 2 *On non-game days the information structure produced will exhibit no bias. On game days, if any information is purchased, the information structure will generically be biased, and the bias will be increasing in H, L, and ρ.*

Proof By construction, on non-home game days, the consumer has a preference for unbiased signals, and so if information is provided it will be unbiased.

Recall that our measure of bias is $\mathbb{B} = p - q$. On game days, if the consumer is willing to purchase a signal for positive amounts, then the firm will provide her the optimal feasible signal structure (p^*, q^*) because this maximizes the price at which the information can be sold, and will charge willingness to pay, $r = p^*\rho H + (1 - q^*)(1 - \rho)L - \bar{u}$. In general the optimal signal structure differs on game days because payoffs to actions (e.g., H and L) differ. So we expect game-day bias, and comparative statics from Proposition 1 pertain. Thus, an increase in H, in L, or in ρ, leads to an increase in p^* and a decrease in q^*, and thus to an increase in $\mathbb{B} = p^* - q^*$.

$\square$

Proposition 3 *When bias is supply-driven bias, if information is purchased on non-game days, the information structure produced will exhibit no bias. On game days, if any information is purchased, the information structure will always be positively biased and the bias will be decreasing in H, L, and ρ.*

Proof As in Proposition 2, by construction, on non-home game days the consumer has a preference for unbiased signals, and so if information is provided it will be unbiased.

As for game days, the firm's problem is to maximize it's payoff. We can write this maximization problem as

$$\max_{p,q}\{b(p, 1-q, H, L, \rho) - c(p-q)\}, \tag{2.11}$$

subject to the technology constraint,

$$q \leq \phi(p). \tag{2.12}$$

If the technology constraint is binding, the following first order conditions characterize the resulting optimal signal structure, (p^*, q^*) :

$$b_1(p^*, 1-q^*, H, L, \rho) - c'(p^* - q^*) + \lambda\phi'(p^*) = 0$$

and

$$-b_2(p^*, 1-q^*, H, L, \rho) + c'(p^* - q^*) - \lambda = 0.$$

In turn, these first order conditions imply that

$$b_1(p^*, 1-q^*, H, L, \rho) - c'(p^* - q^*) = \phi'(p^*)[b_2(p^*, 1-q^*, H, L, \rho) - c'(p^* - q^*)],$$

which in turn can be rewritten,

$$b_1(p^*, 1-q^*, H, L, \rho) - \phi'(p^*)b_2(p^*, 1-q^*, H, L, \rho) = [1-\phi'(p^*)]c'(p^* - q^*). \tag{2.13}$$

Note that in this last equation each of these terms is positive: $b_1(p^*, 1-q^*, H, L, \rho)$, $-\phi'(p^*)$, $b_2(p^*, 1-q^*, H, L, \rho)$, $[1-\phi'(p^*)]$, and $c'(p^* - q^*)$.

Now we turn to properties of the maximum.

The first property concerns *positive bias.* From inspection of the firm's payoff (7), it is clear that the firm will always choose a positively-biased signal structure, i.e., set $p \geq q$. To see this, consider some signal structure that is not positively biased. If $p < q$ then the firm could instead choose an alternative structure, $p' = q$ and $q' = p$, and $c(p-q)$ would be the same, but $b(p, 1-q, H, L, \rho)$ would be strictly larger.

Next, we need to consider two possibilities for the maximum — the optimum is on the the interior of the feasible set, $\Phi(p, q)$, or on boundary. We start with the latter case.

Suppose the maximum is on a boundary. It cannot be on the lower boundary ($p+q=1$) because if it is, it has the same properties, from the consumer's perspective, as $p=q=0.5$ (as is clear from posteriors, (1) and (2) in the text). But as we have seen, it is always optimal for the firm to introduce a small amount of bias. Thus, an optimum on the boundary of $\Phi(p,q)$ must lie on the technology constraint, so (8) is binding. We have already seen that solution in this case is characterized by (9).

The comparative statics exercises are standard. Consider, for example, an increase in H. Since we are on a boundary, we substitute $q^* = \phi(p^*)$ into (9), and then evaluate that expression using the implicit function theorem. This gives

$$\frac{\partial p^*}{\partial H} = -\frac{b_{11} - \phi'(p^*)b_{12} - \phi''(p^*)b_2 - \phi'(p^*)b_{21} + \phi'(p^*)^2 b_{22} + \phi''(p^*)c'(\cdot) - [1-\phi'(p^*)]^2 c''(\cdot)}{b_{13} - \phi'(p^*)b_{23}},$$

which is negative (given assumptions about the signs of cross partials and the convexity of $c(\cdot)$. Similar exercises lead to the conclusions,

$$\frac{\partial p^*}{\partial H} < 0, \quad \frac{\partial p^*}{\partial L} < 0, \quad \text{and} \quad \frac{\partial p^*}{\partial \rho} < 0, \tag{2.14}$$

and q^* must in each case move the opposite direction. This gives us the desired results concerning bias, $\mathbb{B} = p^* - q^*$.

Given our assumptions, second order conditions pertain, and the optimum on the technology constraint must be unique. We note, for completeness sake, that there cannot be a corner solution on the technology constraint given our assumptions that both $\phi'(p)$ and $\phi''(p)$ are strictly negative.

Suppose that instead the the optimum is in the interior of the feasible set. In this instance first order conditions for (7) simplify to

$$b_1(p^*, 1-q^*, H, L, \rho) - c'(p^* - q^*) = 0,$$

and

$$-b_2(p^*, 1-q^*, H, L, \rho) + c'(p^* - q^*) = 0.$$

Here, comparative static exercises analogous to those just presented lead us to the same conclusion just presented.

$\square$

References

Artstein-Avidan, S. and Dillenberger, D., 2010. “Dynamic Disappointment Aversion: Don’t Tell Me Anything Until You Know For Sure,” PIER Working Paper Archive, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Baron, D.P., 2006. “Persistent Media Bias,” *Journal of Public Economics*, 90(1), 1–36.

Brunnermeier, M.K. and Parker, J.A., 2005. “Optimal Expectations,” *American Economic Review*, 95(4).

Caplin, A. and Leahy, J., 2010. “Psychological Expected Utility Theory and Anticipatory Feelings,” *Quarterly Journal of Economics*, 116(1), 55–79.

DellaVigna, S. and Gentzkow, M., 2010. “Persuasion: Empirical Evidence,” *Annual Review of Economics*, 2(1), 643–669.

DellaVigna, S. and Kaplan, E., 2007. “The Fox News Effect: Media Bias and Voting,” *Quarterly Journal of Economics,* 122(3), 1187–1234.

Durante, R. and Knight, B. 2012. “Partisan Control, Media Bias, and Viewer Responses: Evidence from Berlusconi’s Italy,” *Journal of the European Economic Association*, 10(3), 451–481.

Eliaz, K. and Spiegler, R., 2006. “Can Anticipatory Feelings Explain Anomalous Choices of Information Sources?” *Games and Economic Behavior*, 56(1), 87–104.

Epstein, L.G. and Zin, S.E., 1989. “Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework,” *Econometrica,* 937–969.

Gentzkow, M. and Kamenica, E., 2011. “Bayesian Persuasion,” *American Economic Review*, 101(6).

Gentzkow, M., Petek, M., Shapiro, J., and Sinkinson M. 2012. “Do Newspapers Serve the State? Incumbent Party Influence on the US Press, 1869-1928.”

Gentzkow, M.A. and Shapiro, J.M., 2004. "Media, Education and Anti-Americanism in the Muslim World," *Journal of Economic Perspectives*, 18, 117–134.

Gentzkow, M.A. and Shapiro, J.M., 2006. "Media Bias and Reputation," *Journal of Political Economy*, 116, 280–316.

Gentzkow, M.A. and Shapiro, J.M., 2008. "Competition and Truth in the Market for News," *Journal of Economic Perspectives*, 22(2), 133–154.

Gentzkow, M.A. and Shapiro, J.M., 2010. "What Drives Media Slant? Evidence From US Daily Newspapers," *Econometrica,* 78(1), 35–71.

Gentzkow, M. and Shapiro, J., 2011. "Ideological Segregation Online and Offline," *Quarterly Journal of Economics*, 126(4).

Gentzkow, M., Shapiro, J., and Sinkinson M., 2010. "Competition and Ideological Diversity: Historical Evidence from US Newspapers."

Gentzkow, M., Shapiro, J., and Sinkinson, M., 2011. "The Effect of Newspaper Entry and Exit on Electoral Politics," *American Economic Review*, 101(7).

Grant, S. and Kajii, A. and Polak, B., 1998. "Intrinsic Preference for Information," *Journal of Economic Theory*, 83(2), 233-259.

Groseclose, T. and Milyo, J., 2005. "A Measure of Media Bias," *Quarterly Journal of Economics*, 120(4), 1191–1237.

Koszegi, B. and Rabin, M., 2009. "Reference-Dependent Consumption Plans," *American Economic Review*, 99(3), 909–936.

Kreps, D.M. and Porteus, E.L., 1978. "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica,* 185–200.

Mullainathan, S. and Shleifer, A., 2005. "The Market for News," *American Economic Review*, 1031–1053.

Nebeker, F., 1995. *Calculating the Weather: Meteorology in the 20th Century*, Academic Press.

Oster, E., Shoulson, I., and Dorsey, E., 2011. "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease," National Bureau of Economic Research.

Proquest. 1851–2007. *New York Times.*

Proquest. 1889–1993. *Wall Street Journal.*

Segal, U. and Spivak, A., 1990. "First Order versus Second Order Risk Aversion," *Journal of Economic Theory,* 51(1), 111–125.

Sports Reference LLC, 2013. *Baseball Reference.*

Suen, W., 2004. "The Self-Perpetuation of Biased Beliefs," *Economic Journal*, 114(495), 377–396.

CHAPTER III

Behavioral Mechanism Design: Evidence from the Modified First-Price Auctions[1]

3.1 Introduction

Mechanism design is the modern economic analysis of institutions and markets. It has changed the way economists think about optimal institutions when governments are unaware of individual preferences. In the mechanism design literature, the focus is on designing optimal mechanisms (Jackson 2001; Jackson 2003). For example, designing the optimal mechanism is an extremely important problem in the auction literature, where the objective is to increase efficiency by ensuring the object goes to the bidder with the highest value. Experimental research has been helpful in testing whether constructed mechanisms generate the predicted outcomes (Chen and Ledyard 2008; McFadden 2009).[2]

The aim of this paper is to investigate whether seemingly equivalent mechanisms generate the same outcomes in a laboratory setting. Discovering of this issue is crucial, since the mechanism design theory is a standard toolkit for many economists and has affected virtually all areas of policy including regulation, auctions, and environmental policy. An important finding in this literature is that theoretically outcome-equivalent mechanisms do not necessarily lead to the same outcome (Kagel 1995; Kagel and Levin 2011). Indeed, the auction literature provided remarkable evidence

[1]We are indebted to Andrew Schotter for his guidance and help throughout the project. This project would never been possible without his wisdom and support. We also thank Alberto Bisin, Yan Chen, Benjamin Chiao, Syngjoo Choi, James C. Cox, Matthew Embrey, Guillaume R. Frechette, Basak Gunes, Charles Holt, Eric S. Maskin, Daisuke Nakajima, Erkut Ozbay, Charles R. Plott, Debraj Ray, Alejandro Solis and Viktor Tsyrennikov for very helpful comments. This research was supported by the C.V. Starr Center for Applied Economics and the Center for Experimental Social Science of NYU and Department of Economics at University of Michigan. Any remaining errors are ours. Emails: yusufcan@umich.edu, neslihan@umich.edu, setay@umich.edu

[2]In particular, experiments were helpful in testing auctions (Kagel 1995; List 2003; Filiz-Ozbay and Ozbay 2007; Kagel and Levin 2009), optimal contracting (Healy, Ledyard, Noussair, Thronson, Ulrich and Varsi 2007), prediction markets (Healy, Ledyard, Linardi, and Lowery 2010), matching (Niederle, Roth and Sonmez 2008) and public goods (Attiyeh, Franciosi, Isaac and James 2000; Kawagoe and Mori 2001; Chen 2004; Healy 2006).

that even strategically equivalent mechanisms do not generate equal revenues: the first-price sealed-bid and Dutch auctions perform differently in the laboratory settings (Coppinger, Smith and Titus 1980; Cox, Roberson and Smith 1982; Turocy, Watson and Battalio 2007; Katok and Kwasnica 2008).[3] Note that strategic equivalence is stronger than outcome-equivalence since for every strategy in the Dutch auction, there is another strategy in the first-price auction which results in the same outcome; hence one could be tempted to conclude that both auction formats are identical.

While these results are initially striking, one might argue that these are fundamentally different institutions. They differ in the way that they are implemented: the Dutch auction is progressive and open, while the first-price auction is static and sealed (Vickrey 1961; Krishna 2002). Indeed, these procedural differences are vital for an agent whose risk preference is outside of the standard expected utility model. Even though these institutions are strategically-equivalent, the revenue equivalence result holds only when the agents are expected utility maximizers (Karni 1988).[4]

The novelty of this paper is to design and experimentally contrast mechanisms that share exactly the same structure/procedure *and* are strategically-equivalent. In contrast to earlier studies, the risk preferences of agents are not important in this class of mechanisms. We consider an auction environment with a single indivisible object and buyers with independent private values. We introduce a class of mechanisms, so called Mechanism (α), that generalizes the standard first-price sealed bid auction. In Mechanism (α), buyers are asked to submit a value which will then be multiplied by α to calculate the bids in the auction. As in the first-price sealed-bid auction, the buyer with the highest bid gets the good and pays her/his bid. Note that, in equilibrium, the "calculated bids" in the Mechanism (α) are identical for any α. In addition, Mechanism (α) has the same structure under different α's with a small change in how the outcome function translates the values to the bids. Therefore, this family of mechanisms are outcome-equivalent without the need for particular assumptions, such as on risk preferences.

We use the first-price sealed-bid auction as our base mechanism. There are two reasons for this choice. First, the first-price sealed-bid auction is commonly used both in the mechanism design literature and in the real world. Second, we can construct an environment with exactly the same structure across auctions. As a further control to keep the environment same, we provide a "bid calculator" to the subjects so that in all mechanisms calculating the bids (transformations of submitted values to bids) does not create any additional complexity to the auctions than would be inherited by the standard first-price auction. This is in line with Jackson (2001) who argues that complexity

[3] Lucking-Reiley (1999) shows that these auctions are not revenue equivalent in a field setting.

[4] For both the first-price and Dutch auctions, Nakajima (2011) characterizes an equilibrium when bidders have non-expected utility preferences, particularly exhibiting the Allais paradox.

may explain why some outcome-equivalent mechanisms do not perform the same way.

In order to address the question of whether individuals behave equivalently in these strategically-equivalent mechanisms, we run an experiment considering three treatments: Mechanism (0.9), Mechanism (1), and Mechanism (1.1), where Mechanism (1) corresponds to the standard first-price sealed-bid auction. We pick α's close to 1 on purpose to keep the mechanisms as similar as possible. In addition, we pick α's symmetrically around 1 so that the behavioral differences will not be attributed to the distance from the standard first-price auction.

In contrast with what theory predicts, we do not observe outcome-equivalence. We establish the following conclusions: (i) we find that the subjects bidding behavior differed most notably between Mechanism (0.9) and Mechanism (1), where those participating in Mechanism (0.9) shaved their bids by approximately 5%, and (ii) the revenue equivalence does not hold; the standard first-price auction generated higher revenue than the Mechanism (0.9)–subjects who participated in Mechanism (0.9) earned approximately 20% more compared to Mechanism (1).

Our findings provide new challenges for the auction theory as well as the mechanism design theory in general. In order to design optimal mechanisms, one needs to consider the behavioral aspects of mechanisms. The behavioral anomalies might then be taken into account to provide better theories and more efficient mechanisms. The next section describes the Mechanisms (α) and shows the correspondence between the first-price sealed-bid auction. In Section 3, experimental design and procedures are explained. Section 4 presents the data analysis. Section 5 concludes.

3.2 Mechanism (α)

There is a single indivisible good for sale and there are N bidders with private valuations, v_i. The private values are distributed uniformly over the range $V = [0, 100]$. In Mechanism (α), each potential buyer is asked to submit a sealed-value s_i instead of a sealed-bid. The submitted-values are multiplied by α to calculate the bids in the auction. The potential buyer with the highest calculated-bid gets the good and pays her calculated-bid.

It is easy to see that this family of mechanisms are identical and outcome-equivalent independent of bidders' risk preferences. Assume that there exists an equilibrium in the standard first-price auction ($\alpha = 1$). For any $\alpha > 0$, $\frac{1}{\alpha}$ times the original strategy will generate an equilibrium with the same outcome in Mechanism (α). In other words, if there is no incentive to unilaterally deviate from the equilibrium when $\alpha = 1$, then no one would gain by deviating unilaterally from the described equilibrium in Mechanism (α).

Now, for the illustration purposes, we derive the unique symmetric equilibrium strategy when bidders are assumed to have the same von-Neumann-Morgenstern utility function $u(.)$ with $u(0) = 0$, $u^{'} > 0$ and $u^{''} \leq 0$.[5] Given $\alpha > 0$, assume all other players $j \neq i$ follow symmetric, increasing and differentiable equilibrium strategy $S_\alpha : V \to V$. Bidder i is facing a trade off between the winning probability and the gain of winning. Bidder i's maximization problem is:

$$\max_s \; u(v_i - \alpha s) \left[\frac{S_\alpha^{-1}(s)}{100} \right]^{N-1}$$

At a symmetric equilibrium, $s = S_\alpha(v_i)$. Together with the first order condition, this gives

$$(N-1) \frac{u(v_i - \alpha S_\alpha(v_i))}{u^{'}(v_i - \alpha S_\alpha(v_i))} = \alpha S_\alpha^{\prime}(v_i) v_i \tag{3.1}$$

where $S_\alpha(0) = 0$.

If agents have Constant Relative Risk Aversion (CRRA) utility functions with risk aversion coefficient equal to $1 - r$, the unique symmetric equilibrium strategy is:

$$S_\alpha(v_i) = \frac{N-1}{\alpha(N-1+r)} v_i. \tag{3.2}$$

In equilibrium, the submitted-value function, given by equation (crravalue), multiplied with α gives the calculated-bid function. Therefore, it is easy to see that the bids are no longer a function of α and therefore, Mechanism (α) generates the same bid function regardless of the α. That is,

$$\alpha S_\alpha(v_i) = \frac{N-1}{(N-1+r)} v_i.$$

In addition, note that, $\alpha = 1$ corresponds to the standard first-price sealed-bid auction.

3.3 Experimental Procedure

The experiments were performed in the RCGD experimental lab at the Institute for Social Research at the University of Michigan during September 2011.[6] Subjects were recruited from the students of the university (through the ORSEE recruitment program which sends an E-mail to all university students who are enrolled to the program).

The experiment consisted of three treatments, each consisting of three sessions. In each session,

[5] Homogeneity assumption simplifies the illustration of the model, however, our results are independent of it, as shown in Appendix B.

[6] Pilot experiments were run at New York University at the CESS Laboratory.

there were 8 to 16 subjects and 20 rounds. In each round, groups of 4 subjects were formed randomly. Each subject participated in only one of the sessions. The sessions took approximately one hour. Subjects earned laboratory currency (points) which was then converted into cash at the end of the session. A conversion rate of .30 cents per point earned was used.

The treatments are shown in the table below.

	# subjects	# subjects in each auction	# rounds	α
Mechanism (1)	36	4	20	1.0
Mechanism (0.9)	32	4	20	0.9
Mechanism (1.1)	36	4	20	1.1

Table 3.1: Experimental Design

In the three treatments, we chose mechanisms with different α's corresponding to the different treatments. Notice that the Mechanism (1) is simply a first-price sealed-bid auction. Mechanism (0.9) and Mechanism (1.1) are both theoretically equivalent to Mechanism (1), but we would like to see whether they are behaviorally equivalent as well.

In our experiment, participants were seated individually in visually isolated cubicles with computers in front of them. Then, they received instructions on the computer screen (see Appendix A). Instructions were also read aloud in order to make sure that the information was common knowledge. We follow a between subjects design, so each subject participate in only one of the treatments. Subjects were told that there would be 20 rounds and each period, they would be randomly re-grouped with 3 other people in the room without knowing the identities of these people. In each period the value of the object for each subject was determined by a random number generator program in front of them. The values were between 1 and 100 where every number of two decimal places was equally likely.[7] At each period, subjects viewed a screen that showed their value and contained three sections: a "bid calculator" for testing submitted values,[8] a "value submission" section where submitted values were to be entered, and a history sheet indicating results from previous rounds. The subject with the highest calculated bid, which was equal to the submitted value multiplied with the α corresponding to that treatment, won the object in his/her group. The profit of the winner was determined by the difference between his/her value and the calculated bid. In all treatments, the winner was determined randomly with equal probability in case of a tie. At the end of each round, subjects could view the outcome of the auction (whether they won and their profit in points)

[7]We allowed them to enter decimal places to enable a continuous strategy space.

[8]A bid calculator calculates what would be the bid of the subject for a given submitted value.

and review their submitted value and calculated bid.

The main aim of this paper is to see if there is any behavioral difference between these theoretically equivalent mechanisms. We designed the experiment in such a way that the mechanisms are not complicated. Some may argue that any behavioral difference may be attributed to the difference in the complexity of the mechanisms, i.e., the reason for different outcomes could be simply because the subjects failed to recognize the equilibrium of the new game due to its complexity; however, complications of this sort are prevented with the use of the "bid calculator." Also, notice that the Mechanisms (0.9) and (1.1) are not complicated in the sense that if one can solve the equilibrium in the first-price sealed-bid auction then one can solve it in these mechanisms as well.[9]

3.4 Results

As we have demonstrated earlier, for any α_1 and α_2 not equal to zero, the Mechanisms (α_1) and (α_2) are theoretically equivalent. However, if the mechanisms are not behaviorally equivalent, then this may open up a new dimension for designing mechanisms.

3.4.1 Bidding Behavior in the Mechanisms

If individuals have CRRA utility functions with risk aversion coefficient equal to $1 - r$, equation (3.2) implies linear bidding behavior. In order to test for linearity, we run the following regression for each treatment:[10]

$$Bid_i = \beta_0 + \beta_1 value_i + \beta_2 value_i^2 + u_i.$$

We cannot reject the null hypothesis that β_2 equals to zero in all treatments at 5% level, which provides support for linearity of the bid function.[11]

Next, for each mechanism, we regress bids on values and a constant term (see Table 2); we find that the slope (robust standard error) of the estimated bid function of the Mechanism (1) is 0.886 (0.016). Constant term (robust standard error) is equal to -0.225 (0.513). However, the constant term is not significant at the 5% level. If we repeat the regression without a constant term, the slope coefficient (robust standard error) is 0.882 (0.008).

[9] We do not expect Mechanisms (0.9) and (1.1) to have different levels of complexity even if one argues Mechanism (1) has less cognitive load.

[10] Throughout the paper we cluster observations at the session level.

[11] The associated p-value for β_2 is 0.277 for Mechanism (1), 0.223 for Mechanism (0.9), and 0.202 for Mechanism (1.1).

We observe that subjects are shaving their bids approximately 5% more in Mechanism (0.9). Indeed, we find that, in the Mechanism (0.9), the coefficient (robust standard error) of the estimated bid function is 0.829 (0.011). For Mechanism (1.1), the observed behavior is similar to Mechanism (1). The coefficient (robust standard error) of the estimated bid function is 0.874 (0.010). To make the comparisons easier, Table 2 summarizes the estimated bid functions, both with and without constants.

	With constant		Without constant	No. of
	constant	value	value	observations
Mechanism (1)	-0.225	0.886***	0.882***	720
	(0.513)	(0.016)	(0.008)	
Mechanism (0.9)	0.215	0.829***	0.832***	640
	(0.539)	(0.011)	(0.006)	
Mechanism (1.1)	-0.615	0.874***	0.865***	720
	(0.475)	(0.010)	(0.006)	

Standard errors are in parenthesis, *** 1% significance level.

Table 3.2: Estimated Bid Functions

3.4.2 Are differences significant?

Although the Mechanism (0.9) and the Mechanism (1) are TOE, our data suggests that there is a behavioral difference between mechanisms. In order to test for the equivalence of the two bid functions, we have used the dummy variable approach. We run the following regression:

$$Bid_i = \beta_0 + \beta_1 D_i + \beta_2 value_i + \beta_3 value_i * D_i + u_i$$

where D_i is equal to 1 if data point is coming from the Mechanism (0.9), 0 otherwise. We reject that the two bid functions have the same slope coefficients at 5% level ($p - value = 0.022$). We also repeat the same regression without a constant and we still conclude that the interaction term is significantly different than zero (p-value=0.008). Therefore, we reject the null hypothesis that the estimated bid functions are the same.

Pair-wise comparisons of the regressions show that the estimated bid function of the Mechanism (1.1) is significantly different (at the 5% significance level) than the Mechanism (0.9), but is not significantly different from the Mechanism (1.0). Table 3 documents these findings.

The most dramatic difference across the mechanisms is seen in the earnings made by those subjects participating in Mechanism (0.9) compared to Mechanism (1). Those who participated in

	With constant		Without constant	No. of
	β_1	β_3	β_3	observations
Mechanism (1) to (0.9)	0.441	-0.057**	-0.050***	1360
	(0.666)	(0.017)	(0.009)	
Mechanism (1) to (1.1)	-0.390	-0.012	-0.018	1440
	(0.626)	(0.017)	(0.009)	
Mechanism (0.9) to (1.1)	-0.830	0.045**	0.032***	1360
	(0.643)	(0.014)	(0.008)	

Standard errors are in parenthesis, ***1% significance level and **5% significance level.

Table 3.3: Pair-wise Comparison of Estimated Bid Functions

the former treatment made approximately 20% more in earnings, as seen by comparing the average earnings in points of 39.73 in the Mechanism (0.9) to 32.47 in the Mechanism (1).[12] Of course, this intriguing difference indicates a revenue difference across the mechanisms where one would have expected equivalence; this empirical anomaly suggests that one should be careful in picking the optimal mechanism even among the theoretically equivalent class of mechanisms.

3.4.3 Time Trends

Testing for adjustments in bidding over time for each mechanism is not only interesting but also necessary in order to conclude that these mechanisms differ. It may be the case that bidding behavior in Mechanism (0.9) is getting similar to the bidding behavior in the other mechanisms over time. We therefore repeat our regression analysis by adding one more explanatory variable, "iro," which is equal to the inverse of round. A nonlinear adjustment process is preferred over a linear adjustment process, since this allows for a rapid learning in the first rounds.[13] In any case, results do not depend on this specification. We get the same results if we instead add "round."

Table 3 reports the coefficients of the regressions. The time trend coefficient is not significant in the Mechanism (1.1). In the Mechanism (0.9) and (1.0), the time trend coefficient is negative and significant at the 5% significance level. The negative time trend coefficient suggests higher bidding over time. However, we see that even if bids are increasing in Mechanism (0.9), subjects do not bid as aggressively as in Mechanism (1).[14]

[12]The average earnings of the Mechanism (1.1) was 35.22, which is in line with the fact that the estimated bid function for subjects participating in this treatment was slightly less than those participating in Mechanism (1) but higher than those participating in Mechanism (0.9).

[13]See Kagel (1995, pages 521-523) for a discussion on this issue.

[14]To see this compare the coefficients of *value* in the regressions: 0.909 with 0.859.

	Value	Iro	No. of observations
Mechanism (1)	0.909***	-10.632**	720
	(0.007)	(1.100)	
Mechanism (0.9)	0.859***	-9.536**	640
	(0.009)	(1.026)	
Mechanism (1.1)	0.886***	-8.046	720
	(0.005)	(2.839)	

Standard errors are in parenthesis, *** 1% significance level.

Table 3.4: Time Trends

3.4.4 Can Heterogeneity be an Explanation?

As we have already shown, CRRA model itself is not enough to explain the different bids in different mechanisms. The risk aversion coefficient implied by Mechanism (1.0), $1 - r = 0.60$, is not consistent with the risk preferences observed in the other two treatments where $1 - r = 0.39$ for $\alpha = 0.9$ and $1 - r = 0.53$ for $\alpha = 1.1$.

Bidding theory has been extended to agents with heterogeneous risk preferences (see Cox, Smith and Walker (1982, 1983, 1985, 1988)). The heterogeneous constant relative risk aversion model (CRRAM) assumes that each bidder has a different risk aversion coefficient which is drawn from some distribution Φ on (0,1]. Each bidder is assumed to know only his/her own risk aversion parameter and that other bidders' risk aversion parameters are randomly drawn from the distribution Φ. CRRAM model can explain the overbidding behavior observed in the first-price sealed-bid auctions and the heterogeneity between the agents.

In Appendix B, we show that, for any α, each individual i should bid the same way under each Mechanism (α). However, CRRAM model allows for different bid functions for different αs if there are systematic differences in risk preferences across treatments. Therefore, if subject pools are driven from the same population, then Mechanism (α) generates the same bid function for any α. Therefore, in this section, we would like to test whether there is any evidence of systematic differences in risk preferences across treatments. If subjects in the Mechanism (0.9) are systematically less risk averse, then CRRAM model would explain the differences in the bidding behavior.

In part 2 of our experiment, we elicit individuals' risk preferences. The subjects were presented with 15 situations, each of which introduced the choice between a fixed payoff of a specific amount (the "safe choice") or a 50-50 lottery between a payoff of \$4.00 and of \$0.00. When the subjects submitted all of their choices, the computer randomly selected a situation for each subject, and they received the payoff from whichever option they selected from that situation. The average number of safe choices (robust standard deviation) for each treatment is 5.694 (2.571), 5.688 (2.431) and 5.444 (2.567) out of 15 possible choices for Mechanisms 1.0, 0.9 and 1.1, respectively.

We use the non-parametric Mann-Whitney (Wilcoxon Rank Sum) test for each possible pair of treatments in order to test the hypothesis that the number of safe choices are the same against the alternative hypothesis that one treatment has systematically larger values. Table 4 presents the results. Since we cannot reject the hypothesis that the number of safe choices are the same across any pairwise comparisons, we can conclude that the behavioral differences cannot be consistently explained by the CRRAM model.[15]

	Nonparametric Regression Results
Mechanism (1) to (0.9)	-0.299 (0.765)
Mechanism (1) to (1.1)	0.591 (0.555)
Mechanism (0.9) to (1.1)	0.209 (0.835)

Note: Mann-Whitney U-test (one-tail) based on ranks is used. The null hypothesis is that two sets of coefficients come from the same distribution. The numbers in the cells are the z-statistics and the p-values are given in the brackets.

Table 3.5: Pair-wise Comparison of Risk Aversion Coefficients

3.5 Conclusion

We study the behavioral difference between strategically-equivalent mechanisms that share the exact same environment. In order to do this, we construct the Mechanism (α), where the bid functions are identical across different α's. In general, we observe differences across the mechanisms in terms of estimated bid functions: Mechanism (0.9) differs significantly compared to Mechanism (1) and to Mechanism (1.1). Therefore we also see that revenue equivalence between the mechanisms does not hold.

While we cannot pin down what might cause this significant difference, we conjecture that this might be explained by anchoring and adjustment heuristic (Slovic and Lichtenstein 1971; Tversky and Kahneman 1974).[16] We observe that, for $\alpha = 0.9$, individuals do not adjust their bids sufficiently.

[15] A Kruskal-Wallis test also confirms the same result ($p - value = 0.848$). In addition, we checked whether the composition of subjects in different treatments differ in terms of age, gender, participating in an auction before and the frequency of gambling, which may all affect how individuals bid in auctions, and we do not find any significant differences (Kruskal-Wallis tests; p-values are 0.288, 0.164, 0.574, and 0.220, respectively).

[16] Anchoring and adjustment is defined as different starting points yielding different actions, since adjustments are typically insufficient.

In our experiment, on average, individuals submit 88% of their values in Mechanism (1). If they also submit the 88% of their values in Mechanism (0.9), this corresponds to a bid function with a slope of 0.79. Instead we see that individuals bid 83% of their values. Clearly, individuals make an adjustment but the adjustment is not enough. The reason we do not see a similar insufficient adjustment in Mechanism (1.1), which would imply a higher bid function, might simply be due to the small margin of earnings in that region. Individuals realize insufficient adjustment corresponds to very little earnings, so the adjustment is complete. We see this to our advantage since this shows that mechanisms are not different in their complexity, i.e., subjects are able to make full adjustments.

These findings open up a new research agenda for the mechanism design theory. We provide strong evidence that the behavior of individuals is different even under strategically-equivalent mechanisms that are also procedurally the same. This is suggestive that while searching for the optimal mechanism, theory should incorporate the behavioral aspects of mechanisms. As such, these findings have overriding consequences for the parts of theory which relate to implementation and equivalence, including the Revelation Principle.

3.6 Instructions for the Mechanism (0.9)

Instructions

Thank you for agreeing to participate in this experiment. Please make sure your mobile phones are turned off to avoid interruptions during the meeting. This is an experiment in the economics of decision making. Your participation in this experiment is voluntary but we hope that you will also find the experiment interesting. You will be compensated for your participation. You will be given $7 for showing up. You can earn an additional amount of cash which will be distributed at the end of the experiment. The instructions are simple and you will benefit from following them carefully.

Part 1

In the first part of the experiment there will be a series of auctions. In each round you will participate in an auction with three other participants, for a total of 4 people. Between rounds the people in your group will change randomly. However, you will not know the identities of these people. There will be 20 rounds. In each round, your earnings will depend of your choice and the choices made by the 3 other people in your group. We use points to reward you. At the end of the experiment we will pay you 30 cents for each point you won.

In each round, a fictitious good will be auctioned and each of you will have different values for this good. Each round, your value for the good will be determined by a random number generator. The number will be between 1 and 100 where every number (of two decimal places) is equally likely.

At the beginning of Round 1 you will be shown your value for the good. You will participate in an auction for the good, where your final earnings will be the difference between your value and your bid if you win the auction. However, in this auction you will not directly submit a bid. Instead, you will be asked to enter a "submitted value" for the good and a bid will then be calculated for you. You are allowed to enter a "submitted value" that is different from your actual value. The server will collect each participants "submitted value" from Round 1, and your "submitted values" will be multiplied by .9 to determine your calculated bid. For example, say your submitted value is 60, then a calculated bid of $.9*60 = 54$ will be submitted for you. Please note that each round a bid-calculator will be provided for you so you may test out different submitted values before you make your final decision.

The computer randomly forms groups of four participants. Within each group, the calculated bids will be compared. The results of Round 1 will then be displayed. You will be informed whether or not you held the highest calculated bid in your group. Only those of you with the highest calculated bid in their group will get the good, and his/her earnings will be equal to the difference between his/her value and calculated bid. In the case of a tie, the winner will be determined randomly with equal probability. If you are the winner, your earnings will be:

Earnings = Your value for the good - .9 * your submitted value

If the highest calculated bid is not yours (or if you lose in the case of a tie), then you earn nothing in the auction. So, your earnings will be:

Earnings = 0

That will end Round 1, and then Round 2 will begin. The same procedure will be used for all 20 Rounds. After each round you will be able to see whether you have won the auction and your earnings in that round on your computer screen. Your final earnings at the end of the experiment will be the sum of earnings over the 20 rounds. Remember that at the end of the experiment you will receive the show-up fee and your total points/earnings will be multiplied by 30 cents to calculate your final payment.

Part 2

You will now be presented with several Situations. Each Situation will present you with the choice between a Fixed Payoff of a specific amount, or a 50-50 Lottery between a payoff of $4.00 and of $0.00. When you have made all of your choices, the computer will randomly select a Situation, and you will receive the payoff from whichever option you selected. You will then be asked to answer questions from a quick and confidential survey.

Table 3.6: Situations for Risk Elicitation

Situation	Lottery	Fixed Payoff
1	50% chance of $4.00 and 50% chance of $0.00	$0.25
2	50% chance of $4.00 and 50% chance of $0.00	$0.50
3	50% chance of $4.00 and 50% chance of $0.00	$0.75
4	50% chance of $4.00 and 50% chance of $0.00	$1.00
5	50% chance of $4.00 and 50% chance of $0.00	$1.25
6	50% chance of $4.00 and 50% chance of $0.00	$1.50
7	50% chance of $4.00 and 50% chance of $0.00	$1.75
8	50% chance of $4.00 and 50% chance of $0.00	$2.00
9	50% chance of $4.00 and 50% chance of $0.00	$2.25
10	50% chance of $4.00 and 50% chance of $0.00	$2.50
11	50% chance of $4.00 and 50% chance of $0.00	$2.75
12	50% chance of $4.00 and 50% chance of $0.00	$3.00
13	50% chance of $4.00 and 50% chance of $0.00	$3.25
14	50% chance of $4.00 and 50% chance of $0.00	$3.50
15	50% chance of $4.00 and 50% chance of $0.00	$3.75

Survey Questions

Please answer ALL of the questions in this brief survey as accurately as you can. All answers are confidential, and in fact your answers are linked only to your participant ID for today's experiment, and not your name or student ID.

1. What is you age in years? (Enter: Integer.)

2. What is your gender? (Enter: Male, Female.)

3. What is you major? (Enter: String.)

4. What strategy did you use in the auctions? (Enter: String.)

5. Have you ever participated in an auction before? (Enter: Yes, No.)

6. How often have you gambled or purchased lottery tickets in the past year? (Enter: Very frequently, Frequently, Sometimes, Rarely, Never.)

3.7 Can heterogeneity be an explanation?

We now adopt the CRRAM model presented in Cox, Smith and Walker (1988) to include the possibility of heterogeneous risk averse bidders, in which the homogeneous constant relative risk aversion model is a special case. For simplicity, we first solve the bid function for Mechanism (1). There is a single good for sale and there are N bidders with private valuations, v_i. Each bidder has von-Neumann-Morgenstern utility function $u(v_i - b_i, r_i)$, where risk preference r_i is randomly drawn from some distribution function Φ on (0,1] and b_i denotes the bid of agent i. Assume that $u(x, r)$ is twice continuously differentiable and strictly increasing with respect to the first component and $u(0, r) = 0$, for all $r \in (0, 1]$. Also, assume that $u(x, r)$ is strictly log-concave in r, for each $r \in (0, 1]$.

Assume that each bidder believes that his/her rivals will use the bid function $b(v, r)$, which is strictly increasing in v and $b(0, r) = 0$ for all $r \in (0, 1]$. If the bid function has a v-inverse function $\pi(b, r)$, that is differentiable and strictly increasing in b, then the probability of all $n - 1$ rivals of bidder i will bid amounts less than or equal to b is

$$G(b) = [\int_r H(\pi(b, r)) d\Phi(r)]^{n-1} \tag{3.3}$$

where H is the uniform distribution on [0,100].

Therefore, the expected utility of bidding b_i to bidder i is

$$G(b_i)u(v_i - b_i, r_i) \tag{3.4}$$

The first order condition with respect to b_i is

$$G'(b_i)u(v_i - b_i, r_i) - G(b_i)u_1(v_i - b_i, r_i) = 0 \tag{3.5}$$

If $\pi(b, r)$ is the v-inverse of an equilibrium bid function, then it must be a best response for bidder i and, therefore, should satisfy the first order condition.

$$G'(b_i)u(\pi(b_i, r_i) - b_i, r_i) - G(b_i)u_1(\pi(b_i, r_i) - b_i, r_i) = 0 \tag{3.6}$$

This implies

$$\frac{d(G(b_i)u(\pi(b_i, r_i) - b_i, r_i))}{db_i} = G(b_i)u_1(\pi(b_i, r_i) - b_i, r_i)\pi_1(b_i, r_i) \tag{3.7}$$

Integrating (3.7) yields

$$G(b_i)u(\pi(b_i, r_i) - b_i, r_i) = \int_0^{b_i} G(y)u_1(\pi(y, r_i) - y, r_i)\pi_1(y, r_i) \tag{3.8}$$

Cox et al. (1988) show that b_i maximizes bidder i's expected utility, when his/her value is $\pi(b_i, r_i)$, for any $b_i > 0$ in the domain of $\pi(., r_i)$. Hence, $\pi(b, r)$ given by (3.8) is the v-inverse of an equilibrium bid function. Next, they show that $(b_i, \pi(b_i, r_i))$ yields a global maximum of (3.4).

We will now show that, for any α, Mechanism (α) implies the same bid function even when we allow for the possibility of heterogeneous agents. Assume that each bidder believes that his/her rivals will use a submitted value function $s(v, r)$, with v-inverse function $p(s, r)$. The probability that all $n - 1$ rivals of bidder i will submit values less than or equal to s is

$$K(s) = [\int_r H(p(s, r))d\Phi(r)]^{n-1} \tag{3.9}$$

And, the expected utility of reporting a value s_i equals

$$K(s_i)u(v_i - \alpha s_i, r_i) \tag{3.10}$$

The first order condition is

$$K'(s_i)u(v_i - \alpha s_i, r_i) - \alpha K(s_i)u_1(v_i - \alpha s_i, r_i) = 0 \tag{3.11}$$

Now, it is easy to see $s(v_i, r_i) = \frac{b(v_i, r_i)}{\alpha}$ satisfies equation (3.11). First, note that, $p(\frac{b(v_i,r_i)}{\alpha}, r_i) = \pi(b_i, r_i)$ if $s(v, r)$ is the equilibrium submitted value function and, therefore, $K(\frac{b(v_i,r_i)}{\alpha}) = G(b_i)$ and $\frac{1}{\alpha}K'(\frac{b(v_i,r_i)}{\alpha}) = G'(b_i)$. Therefore, for each individual i, the bid function of the Mechanism (1) and the bid functions of the Mechanism (α) are the same.

References

Attiyeh, G., Franciosi, R. and Isaac, R. M. (2000). "Experiments with the Pivot Process for Providing Public Goods." Public Choice. 102, 95-114.

Chen, Y. (2004). "Incentive-Compatible Mechanisms for Public Goods: A Survey of Experimental Research." in C. Plott and V. L. Smith, (eds.), The Handbook of Experimental Economics Results, Amsterdam: Elsevier Press.

Chen, Y. and Ledyard, J. O. (2008). "Mechanism Design Experiments." in S. N. Durlauf and L.E. Blume (eds.), The New Palgrave Dictionary of Economics, Macmillan.

Coppinger, V.M., Smith, V.L., and Titus, J.A. (1980). "Incentives and Behavior in English, Dutch and Sealed-Bid Auctions." Economic Inquiry, 18 (1), 1-22.

Cox, J. C. and Oxaca, R. L. (1996). "Is Bidding Behavior Consistent with Bidding Theory for Private Value Auctions." in R. Mark (ed.), Research in Experimental Economics, vol. 6, Greenwich, CT: JAI Press, 131-148.

Cox, J. C., Roberson B., and Smith, V.L. (1982). "Theory and Behavior of Single Object Auctions." in V. L. Smith (ed.), Research in Experimental Economics, vol. 2, JAI Press.

Cox, J. C., Smith, V. L. and Walker, J.M. (1982). "Auction Market Theory of Heterogeneous Bidders." Economics Letters. 9(4), 319-325.

Cox, J. C., Smith, V. L. and Walker, J.M. (1983). "Tests of a Heterogeneous Bidder's Theory of First Price Auctions." Economics Letters. 12(3-4), 207-212.

Cox, J. C., Smith, V. L. and Walker, J.M. (1985). "Experimental Development of Sealed-Bid Auction Theory; Calibrating Controls for Risk Aversion." American Economics Review. 75, 160-165.

Cox, J. C., Smith, V. L. and Walker, J.M. (1988). "Theory and Individual Behavior of First-Price

Auctions." Journal of Risk and Uncertainty, 1(1), 61-99.

Filiz-Ozbay, E. and Ozbay, E. (2007). "Auctions with Anticipated Regret: Theory and Experiment." American Economic Review, 97(4), 1407-1418.

Healy, P. J. (2006). "Learning dynamics for mechanism design: An experimental comparison of public goods mechanisms." Journal of Economic Theory, 129(1), 114-149.

Healy, P., Ledyard, J., Noussair, C., Thronson H., Ulrich P, and Varsi G. (2007). "Contracting inside an organization: An experimental study." Experimental Economics, 10(2), 143-167.

Healy, P., Linardi, S., Lowery, J.R., Ledyard, J.O. (2010). "Prediction Markets: Alternative Mechanisms for Complex Environments with Few Traders." Management Science, 56(11), 1977-1996.

Isaac R. and James D. (2000). "Robustness of the Incentive Compatible Combinatorial Auction." Experimental Economics, 3(1), 31- 53.

Jackson, M. O. (2001). "A Crash Course in Implementation Theory. Social Choice and Welfare. 18, 654-708.

Jackson, M.O. (2003). "Mechanism Theory. " in U. Derigs (ed.), The Encyclopedia of Life Support Systems, Oxford.

Kagel, J. H. (1995). "Auctions: A Survey of Experimental Research." In J.H. Kagel and A.E. Roth (eds.), The Handbook of Experimental Economics, Princeton, 501-585.

Kagel, J. and Levin, D. (2009). "Implementing Efficient Multi-Object Auction Institutions: An Experimental Study of the Performance of Boundedly Rational Agents." Games and Economic Behavior, vol 66. (1): 221-237

Kagel, J. and Levin, D. (2011). "Auctions: A Survey of Experimental Research, 1995 2010. " to appear in the Handbook of Experimental Economics, Vol. 2.

Karni, E. (1988). “On the equivalence between descending bid auctions and first price sealed bid auctions.” Theory and Decision, 25(3), 211-217.

Katok E. and Kwasnica A.M. (2008). “Time is Money: The effect of clock speed on seller’s revenue in Dutch auctions. ” Experimental Economics, 11, 344-357.

Kawagoe T. and Mori T. (2001). “Can the Pivotal Mechanism Induce Truth-Telling? An Experimental Study.” Public Choice. 108, 331-354.

Krishna, V. (2002). “Auction Theory.” Academic Press.

Lichtenstein, S. and Slovic, P. (1971). “Reversals of Preference Between Bids and Choices in Gambling Situations. Journal of Experimental Psychology, 89, 46-55.

List, J. (2003). “Does Market Experience Eliminate Market Anomalies?” The Quarterly Journal of Economics, 118(1), 41-71.

Lucking-Reiley, D. (1999). “Using Field Experiments to Test Equivalence between Auction Formats: Magic on the Internet” The American Economic Review, 89(5), 1063-1080.

McFadden, D. (2009). “The human side of mechanism design: a tribute to Leo Hurwicz and Jean-Jacque Laffont.” Review of Economic Design, 13 (1-2) 77-100.

Nakajima, D. (2011). “First-price auctions, Dutch auctions, and buy-it-now prices with Allais paradox bidders.” Theoretical Economics, 6(3), 473-498.

Niederle, Muriel, Alvin E. Roth and Tayfun Sonmez (2008). “Matching.” The New Palgrave Dictionary of Economics, 2nd edition, Eds. Steven N. Durlauf and Lawrence E. Blume, Palgrave Macmillan.

Turocy, T.L., Watson E. and Battalio R.C. (2007). “Framing the first-price auction.” Experimental Economics, 10(1), 37- 51.