THREE ESSAYS ON THE ECONOMICS OF THE NONPROFIT SECTOR

by

Daniel Jones

B.A. in Economics, University of Alabama at Birmingham, 2008

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

PhD in Economics

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

THE DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Daniel Jones

It was defended on

May 3, 2013

and approved by

Dr. Lise Vesterlund, University of Pittsburgh, Economics

Dr. Randall Walsh, University of Pittsburgh, Economics

Dr. Werner Troesken, University of Pittsburgh, Economics

Dr. Sera Linardi, University of Pittsburgh, GSPIA

Dissertation Advisors: Dr. Lise Vesterlund, University of Pittsburgh, Economics,

Dr. Randall Walsh, University of Pittsburgh, Economics

THREE ESSAYS ON THE ECONOMICS OF THE NONPROFIT SECTOR

Daniel Jones, PhD

University of Pittsburgh, 2013

In this dissertation, I study the economics of the nonprofit sector. In particular, I focus on individuals' motives to contribute to the nonprofit sector either through charitable giving or employment within the sector.

In the first chapter, I ask: how do state lotteries designed to fund education impact overall funding available for education? I find that -- while state government net expenditures on education do not significantly increase -- charitable donations to education fall. This results in a decrease in funding available for education. The large degree of charitable crowd-out observed here (relative to other papers in the literature) seems to be driven by how aware donors are of the purported beneficiary of the lottery.

Evidence on whether nonprofit workers earn less than for-profit workers is mixed. In the second chapter, I attempt to reconcile mixed results by considering whether we should always expect to see wage gaps even when there are workers who are willing to work for less because they are motivated by the mission of the nonprofit. I argue that we should only expect wage gaps when labor demand of the nonprofit sector of an industry is low. When labor demand is high, there are not enough "motivated" workers to fulfill demand and so nonprofits must raise wages. I find empirical evidence consistent with these predictions. Penalties for working in a nonprofit are largest in areas where nonprofits require a small share of the labor force. In these same locations, the quality of work being done and worker job satisfaction are substantially higher than for-profits.

In the third chapter, Sera Linardi and I ask: how does visibility impact prosocial behavior? We propose that some individuals have "wallflower" preferences and prefer to avoid standing out in a negative way (giving much less than average) *or* a positive way (giving much more than average). Wallflowers, therefore, move towards what they expect the average individual would choose. As a result visibility may only increase giving when donors expect

others' giving to be high. We conduct a laboratory experiment that supports this prediction. "Wallflower" behavior is particularly prominent amongst women.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

This dissertation was shaped by several years' worth of advice and direction from my committee: Lise Vesterlund, Randy Walsh, Sera Linardi and Werner Troesken. From very early on in my time at the University of Pittsburgh, Lise – who co-chaired the committee – pointed me in directions that ultimately led to the chapters in this dissertation. I very much appreciate her guidance, motivation and insight; I also appreciate the financial support she provided to obtain data and conduct experiments throughout my graduate career. Randy – who also co-chaired the committee – has always kept an open door to discuss topics ranging from small methodological issues to much broader thoughts on the direction of my research. Through these many conversations, Randy's advice and support made completing this dissertation much easier. I am grateful for the many hours that Sera devoted to reading drafts of my papers, her excellent and thoughtful feedback, and her positivity and encouragement. I have enjoyed and benefitted from Werner's interest in and willingness to discuss ideas; discussions with Werner at the early stages of the third chapter of this dissertation ("The Supply & Demand of Motivated Labor") motivated me to continue pursuing the idea. Additionally, the opportunity to work as a co-author with Randy and Werner (on research outside of this dissertation) and with Sera (on the third chapter of this dissertation) provided a fantastic education in being an economist. I am grateful for their patience as I learned from them. Finally, I want to thank my parents, Kate and Taylor Jones, for putting me in the position to spend all of my time doing something that I love.

1.0 INTRODUCTION

Humans are, perhaps to an economist at least, surprisingly generous and cooperative. This generosity can be seen immediately in the context of charitable giving; according to the 2009 Giving USA report, even in the midst of a recession, more than $300 billion were contributed to charitable causes in the United States, accounting for roughly 2% of GDP. Elsewhere, individuals appear to be willing to sacrifice wages to work for a nonprofit firm whose mission they support.[1] In countless laboratory experiments in economics and psychology, participants have displayed a willingness to simply give away money to other participants, even in a single anonymous interaction where there is no opportunity for future reciprocity or reputational gains.

This dissertation contributes to a line of research that asks: What motivates individuals to engage in cooperative behaviors or voluntary action for the public good? From the examples above, it seems that individuals do not simply maximize monetary gain, but what else do they maximize? An understanding of such motives is not purely a theoretical curiosity, but instead has important practical consequences for government, employers, fundraisers, and others. As an illustration, consider an individual who makes a donation to a college scholarship fund for underprivileged children in her state. Perhaps she made this donation because she is genuinely concerned about the educational attainment of these children. Or perhaps she made the donation because donating simply makes her feel good, or provides some "warm glow."[2] The full impact of government spending on scholarships for underprivileged children depends critically on which type of motive is prevalent. If the donor gives because of warm glow, she will continue to give when government increases its own spending activities – thereby allowing government to

---

[1] There is some disagreement in the literature about whether nonprofit workers *actually* earn less after controlling for worker characteristics. This is addressed in Chapter 3 of this dissertation.

[2] This "warm glow" motive and its implications for giving were first discussed in landmark papers by Jim Andreoni (1988, 1989).

actually increase funding available for the scholarships. If, on the other hand, the donor is concerned about the amount of funding available for scholarships, government spending serves as a fine substitute for donations and so she may reduce her donations. This, then, would limit the ability of government to impact actual funding available for scholarships. This is just one example of how understanding the motives for generosity can make a critical difference in understanding the economic impact of practical decisions made by governments or other groups. This dissertation studies "doing good" in several settings with the goal of learning about the preferences that lead individuals to do good, which in turn will help us understand the broader impact of government spending, incentives to work for nonprofits, and incentives to donate to nonprofits.

In the second chapter – much like the example discussed above – I consider the impact that the introduction of state lotteries designed to fund education has on overall education funding. Because education relies heavily on charitable contributions, it is necessary to consider the impact of lotteries on both government education spending and on private donations. Using data on state government budgets, donors, and nonprofit firms, I find that earmarking lottery revenues need not increase – and may *decrease* – education funding. First, government treats additional revenue as fungible and directs it to other sources. Second, the announcement of the education lottery appears to decrease donors' willingness to give to education.

This decrease in willingness to give is surprising because much of the literature on the relationship between government activity and charitable donations suggests that donors appear to be relatively unresponsive to government activity.[3] This existing literature *could* be evidence that donors are simply unresponsive to government spending (perhaps because they are motivated by warm glow), but it could also be that donors are largely unaware of the type of government spending often studied in existing work (e.g., government grants to nonprofits). If the latter is true, there remains some uncertainty as to whether donors would respond to government spending if they were aware of it. State lotteries are highly salient because the state heavily advertises the new source of funding for education. They therefore provide the opportunity to re-examine whether crowd-out occurs in an environment where donors are aware of government activity. I provide evidence that higher awareness of the lottery increases crowd-out: states with

[3] Or, where there is a drop in giving, the drop can largely be explained by a reduction in fundraiser effort (Andreoni & Payne, 2003).

higher lottery advertising activities experience larger drops in giving. There are also larger drops in giving in states where citizens vote directly on the introduction of an education lottery as compared to states where the legislature decides. In short, the combination of crowd-out and fungibility suggest that total funding for education may decrease with the introduction of an education lottery. The paper also demonstrates that in this context, when donors are aware of government spending, they respond by decreasing their contributions.

In the third chapter ("The Supply & Demand of Motivated Labor"), I reconsider the well-known -- but not well-understood -- fact that nonprofit workers earn less on average than for-profit workers. Existing empirical work leaves open the question of whether this is driven by a willingness to work for less (the "labor donation hypothesis") or instead simply by the composition of the nonprofit sector. Wage gaps have been found to be present in some industries (e.g., legal services) and absent in other industries (e.g., hospitals). I consider when we should expect labor donations to nonprofits to generate wage gaps and, in doing so, offer an explanation for the inconsistent results in the existing literature.

Unlike much of the existing work, I highlight the importance of nonprofits' labor demand. Specifically, it is only in nonprofit employers' interest to maintain low wages if their labor demand does not exceed the number of workers who are willing to work for less. Otherwise, nonprofits must raise wages to attract other workers. This yields the prediction that wage gaps should be largest when the nonprofit share of labor within an industry is low. As nonprofit share increases, wages should equalize. Using economy-wide Census microdata, I provide evidence consistent with this prediction. Using more detailed data from the nursing home industry which allows me to better control for market conditions and rule out alternative explanations, I also find that the quality of work in nonprofit nursing homes is highest in localities where the wage gap is largest and the nonprofit share low. Finally, to demonstrate that a different "type" of worker sorts into nonprofits when nonprofit share is low, I provide evidence that job satisfaction and the likelihood of engaging in charitable giving (outside of work) are highest amongst nonprofit worker in low nonprofit share areas. Nonprofits and nonprofit workers in higher nonprofit share areas are more similar to for-profits, both with regards to the quality of work being done and the workers' satisfaction and prosocial activities. This provides evidence that the relationship between wage gaps and nonprofit share is driven by "motivated types" sorting into nonprofit jobs. More broadly, this highlights the importance of intrinsic motivation as a factor in individuals' job searches – both for workers and for the quality of work that results from attracting motivated workers. This suggests that increasing pay is not necessarily the clear solution to improve the quality of workers that are attracted to a particular job.

In the fourth chapter ("Wallflowers"), Sera Linardi and I consider the impact of public recognition on giving to charity. Public recognition, of course, is widely used by fundraisers with the goal of increasing contributions. Existing theoretical literature suggests that visibility should increase giving by providing the opportunity to obtain a positive reputation or forcing would-be free-riders to avoid a negative reputation (Benabou & Tirole, 2006). Experimental work has largely found this to be the case, but there are some notable exceptions wherein visibility either has no impact or decreases contributions. This suggests that individuals may not uniformly seek a positive reputation. We propose that some individuals have "wallflower" preferences and avoid both negative *and* positive reputations. To achieve this, wallflowers would choose a contribution close to what they expect that the "average" individual would choose. As a result, visibility should be expected to have a positive impact if wallflowers believe that *others* are making large donations, but could also have a negative impact if wallflowers believe others give little.

We test this idea in a laboratory experiment. Participants are placed in groups of three and can donate money to a real charitable cause. Participants have the opportunity to indicate what donation they would like to choose conditional on every possible combination of group members' contributions. In one treatment (the "Visibility" treatment), contribution decisions and identities are revealed within groups. We predict that, if wallflower preferences are prevalent, participants will respond to visibility by choosing contributions strictly within the range of group members' contributions. We find evidence in support of this prediction. As a result, visibility only has a positive impact when participants believe that the rest of their group has chosen a large contribution. We find that this is especially true of women, both in our laboratory experiment and in a small-scale field experiment.

Taken as a whole, these essays highlight the importance – and complexity – of individuals' motives to engage in prosocial behaviors. Policies or actions that would *seem* to increase public good provision (lotteries, higher wages for nonprofit workers, public recognition in fundraising) are ultimately shown to have substantially more nuanced effects after accounting for individuals' motivations.

2.0 EDUCATION'S GAMBLING PROBLEM: THE IMPACT OF EARMARKING LOTTERY REVENUES FOR EDUCATION ON GOVERNMENT SPENDING AND CHARITABLE GIVING

2.1 INTRODUCTION

Over the past several decades, state governments have come to embrace lotteries as an alternative source of revenue. Lotteries have proven to be successful in this regard; on average, lotteries add nearly 500 million dollars to states' budgets yearly.[4] While a handful of states add lottery revenue to their general funds, states typically earmark the revenue to support particular public goods. States adopt lotteries with the intention of funding causes as diverse as environmental protection, the arts, and support for their elderly, but most commonly lottery funds are earmarked for education. Twenty of the forty-three states that currently sponsor lotteries direct all of their revenues towards education, while several more dedicate at least some fraction to education. However, some existing research suggests that the purported beneficiaries of state lotteries rarely experience a significant increase in state government expenditures (Borg et al., 1991; Erekson et al., 2002; Spindler, 1995)[5].

Even if earmarked lottery revenues do not increase government's contribution to the intended public good, government is of course not the only source of funding for many public goods. In most cases, the causes supported by state lotteries also benefit from and rely on charitable contributions. This is especially true of education. In aggregate, education-related organizations consistently receive more donations than any other secular cause in the United

---

[4] Based on the 2008 *Survey of Government Finances*

[5] This literature is discussed in more detail in section 2.3.

States. Americans donated a total of 38.87 billion dollars towards education in 2011, which is roughly twice the amount of money that was raised through state lotteries in the same year.[6] An examination of government expenditures alone therefore does not capture the full impact that a lottery has on public good provision, as the lottery may also affect charitable contributions.

With this in mind, I examine the impact of the introduction of education lotteries on overall education funding, considering both government expenditures and – as the main focus of the paper – private donations. Standard models of public good provision suggest that if individuals' utilities depend at least in part on the overall level of the public good, then government spending serves as a substitute for charitable contributions (Andreoni, 1989; Bergstrom et al., 1986). Thus, we should expect charitable contributions to decrease with an increase in government spending. Numerous empirical tests have generally found this to be the case, though in many cases the degree of crowd out is relatively small.[7] If donors respond to the *announced* increase in government funding associated with the introduction of a lottery, then this – combined with the fungibility of lottery revenue – may imply that lotteries lead to a *decrease* in total provision.

I assess the degree to which lottery revenue impacts donors' contributions using three individual-level surveys: *Center on Philanthropy Panel Study* (COPPS), *Giving and Volunteering Survey* (GVS), and *Consumer Expenditures Survey* (CES). Collectively, these surveys span from 1989 to 2008, so all of the analysis in the paper focuses on this time period. All of these surveys ask respondents to indicate how much money they have donated recently to a variety of causes, including education. In a difference-in-differences framework I compare the level of education-related donations before and after a state has introduced an education-funding lottery. I find a significant decrease in education giving when an education lottery is introduced.

I then address *why* contributions fall in this context and speak to a more general question in the literature on donors' response to government activity. Andreoni and Payne (2003), Andreoni and Payne (2011) show that the negative relationship between charitable contributions and government grants to nonprofits can in some cases almost entirely be explained by a decrease in fundraiser effort. Their results might suggest that donation decisions are in fact relatively unresponsive to the overall level of the public good. In this paper, I explore a different

---

[6] Giving USA, 2012
[7] See Vesterlund (2006) for a review of the empirical crowd-out literature.

explanation for their result and for the small degree of crowd-out that is often observed in the literature: donors may be largely unaware of government activity in most settings. While this theoretical possibility has been discussed in the literature,[8] to my knowledge this is the first paper to empirically assess the importance of the salience of government spending.

Unlike government spending in the form of grants to nonprofits, the intended increase in spending associated with the introduction of a lottery is highly publicized and the beneficiary is well known. States are eager to advertise that revenues go towards a "good cause," perhaps to overcome moral opposition to the lottery and draw in customers who might not otherwise gamble (Clotfelter & Cook, 1990); advertisements therefore typically include some reminder of the cause supported by lottery revenues (Clotfelter & Cook, 1991). Thus, state lotteries provide the opportunity to test whether donors (and not just fundraisers) respond to government activity in a setting where government spending is highly salient.

To determine whether donors or nonprofits drive crowd-out, I analyze the tax returns of a random sample of nonprofits in the same difference-in-differences framework. I find that an education lottery decreases donations received by education-related organizations by roughly 8%. This is not driven by a change in fundraising behavior. Moreover, there is a negative relationship between donations received and a proxy for a state's lottery advertising expenditures. This suggests that donors' response to (the perception of) increased spending is dependent on the salience of government activity.

A few empirical studies have examined the interaction of lottery expenditures and charitable giving; however, these studies primarily examine the general charitable activities of lottery players. In cross-sectional data, (Borg et al., 1991) report a negative relationship between charitable giving and lottery expenditures; they then provide some evidence that suggests that lottery players would have been contributing less even in the absence of a lottery. Lin and Wu (2007) find a positive relationship between charitable giving and government-sponsored lottery expenditures in Taiwan, where lottery revenues are used to support a variety of public goods.[9] However, in a follow-up paper also based on the Taiwanese lottery, Wu (2012) more fully

---

[8] In particular, both Garrett and Rhine (2010) and Monti (2010) point to higher awareness of government activity as a potentially important difference between direct government spending and spending through government grants. Monti presents a model demonstrating the impact that increased awareness may have on donations.

[9] Devlin (2004) show that there is a positive relationship between charitable donations and participation in smaller scale *charity-sponsored* lottery fundraisers.

accounts for selection bias and shows that there is in fact little relationship between the amount that lottery players spend on the lottery and the amount that they donate to charity. However, the question of whether the introduction of a lottery impacts charitable contributions more generally – and not just at the level of individuals who choose to play the lottery – has been neglected. [10] By focusing on the charitable expenditures of lottery players, these existing results can not speak to the potential for more general crowding-out of donations.

A related theoretical and experimental literature examines the use of lotteries and raffles as substitutes for relying on voluntary contributions. This literature suggests that a fixed-prize lottery can lead to higher public good provision than voluntary contributions (Lange et al., 2007; Morgan, 2000; Morgan & Sefton, 2000). However, these results do not hold for lotteries where the prize is a function of the number of tickets sold (pari-mutuel lotteries). State lotteries include both fixed prize and pari-mutuel components, so I do not claim to speak directly to this theoretical literature here. However, Morgan's model does highlight the idea that donors might view the lottery as an alternative method of contributing to education. If this is the case, a decrease in giving is not necessarily indicative of a decrease in private support for education. I consider this possibility empirically in a later section of this chapter.

The decrease in charitable support for education reported here exacerbates the redistributive concerns associated with lotteries, which are already known to be a highly regressive (albeit voluntary) tax (Grote and Matheson, 2011). In the United States, there is a significantly negative relationship both between (1) income and lottery expenditures (as a fraction of income) and (2) education and lottery expenditures (Clotfelter & Cook, 1991). Charitable giving, on the other hand, shows just the opposite patterns: (1) giving (as a fraction of income) increases with income for most of the range of incomes and (2) giving significantly increases with education (Andreoni, 2006). Thus, lotteries may decrease overall public good provision *and* shift the burden of financing a public good from high- to low-income individuals.

While researchers have examined a variety of issues related to state lotteries, the general impact of state lotteries as a means to finance public goods is not well understood. This paper fills this gap by examining not just government activity but also the impact of lotteries on

[10] One very recent exception is Andreoni, Payne, and Smith's (2013) analysis of donations received by charitable organizations selected to receive funds from the UK National Lottery. They find that, for small organizations, receiving lottery funds actually increases the donations an organization receives. This result is discussed in greater detail in my concluding remarks.

charitable expenditures. In doing so, the results also contribute to the more general literature on the interaction of government activity and charitable giving. Recent work in this area has generally found that crowd-out is largely explained by fundraiser behavior. The results presented here point to the importance of salience of government activity; when donors are more aware of government activity, their behavior is more in line with the crowd-out predicted by classic models of voluntary contributions to public goods.

The remainder of the chapter proceeds as follows. In 2.2, I provide some additional background on education-funding lotteries and charitable donations to education. In section 2.3, I discuss the general empirical strategy used throughout. In section 2.4, I analyze state government finance data; this is done both to test the robustness of existing fungibility results in the time period studied here and also as a means of introducing the lottery "treatment." I then turn to the main focus of the paper: I first examine the impact of the introduction of an education lottery on individuals' education contributions (section 2.5) and then on education-related nonprofit firms' finances (section 2.6). Section 2.7 concludes.

2.2 ADDITIONAL BACKGROUND ON STATE LOTTERIES AND CHARITABLE SUPPORT FOR EDUCATION

Before proceeding to the analysis, some additional detail on state lottery and charitable support for education will help fix ideas. In particular, the degree to which we might expect donors to reduce their contributions depends in part on their perception of the overlap between the causes they support and the specific causes supported by the lottery. Thus, despite the fact that most of the analysis will center on the impact of lotteries on education spending and giving in general (in part because I am unable to decompose education giving any further in the donor-level data), here I discuss which particular causes within education tend to benefit from each source of funding.

As noted, education is typically the most popular secular category of giving in the United States, second only to religious giving. In 2011, 38.87 billion dollars and 13% of all charitable donations went to education-related causes. This figure – and the "education giving" discussed throughout – includes donations to a wide array of education-related organizations: *"giving to the education subsector includes giving to support nonprofit, public, and charter, pre-K through grade 12 schools; non-profit and public colleges and universities; vocational and technical schools; nonprofit and public libraries; education research and policy; adult education programs; tutoring programs; and student services organizations."*[11] However, a majority of donations to education (roughly 78% in 2011) support public and private higher education (including scholarship and financial aid programs). There is a fairly even split between support for public and private institutions: in 2011, private institutions received 55% of donations to higher education, while public institutions received 45% of donations.[12]

So while a broad array of causes fall under the umbrella of "support for education," the main beneficiaries are public and private institutions of higher education. We will see that this is generally true of education lotteries introduced during the sample period as well. In the difference-in-differences framework employed in this paper, the impact of an education lottery is identified by changes that occur in states that introduce a lottery at some point during the sample period. The main donor-level data I use spans from 1989 to 2008 so I focus on this time period throughout. The states that introduced education lotteries during these years and the specific causes that they currently support are listed in Table 2.1.[13] Like private charitable support for education, a majority of these lotteries are currently designed – at least in part – to fund higher education. Many of these lotteries were accompanied by the introduction of large-scale, state-run, lottery-funded scholarship programs.[14] Many of the lottery programs also support programs outside of higher education that often fall within the private nonprofit sector, such as literacy programs and pre-kindergarten programs for low-income children.

---

[11] Source: *Giving USA 2012* report

[12] Source: Council for Aid to Education Annual Survey (2012)

[13] A comprehensive historical listing of specific beneficiaries is not available. All lotteries listed have supported *some* education-related cause(s) since the date indicated. Some states (like South Carolina) adjust the specific composition of their beneficiaries on a year-to-year basis.

[14] Georgia's HOPE Scholarship is a prominent example and seems to have served as a model for several states that followed.

Table 2.1. Education lotteries introduced during sample period

State	Education lottery established	Specific beneficiary**
Georgia	1993	Higher ed. scholarships (public & private schools), funding for pre-K programs
Missouri	1993*	Programs at all levels of public education
New Mexico	1996	Higher ed. scholarships (public universities / community colleges)
Texas	1997*	Public K-12
Vermont	1998*	"Education fund"
Virginia	2000*	Public K-12
Washington	2001*	Higher ed. scholarships/fin. aid (public & private schools), low-income pre-K programs
South Carolina	2002	All levels of education, scholarships/fin. aid for public & private universities
Tennessee	2004	Higher ed. scholarships (public & private), pre-K and after-school programs
Kentucky	2005*	Higher ed. scholarships (public & private), early childhood literacy programs
Oklahoma	2005	Public K-12, Higher ed. grants / loans / scholarships, Other higher ed. programs
North Carolina	2006	Public K-12, Higher ed. scholarships / financial aid, and pre-K programs

* These states already had a lottery (with revenues going towards a different cause or a general fund) but switched to earmarking funds only for education in the year indicated.
** Information on specific beneficiaries is obtained from state lottery websites and is current as of early 2013.

2.3 GENERAL EMPIRICAL STRATEGY

Throughout the paper, I employ a difference-in-differences (DID) approach to identify the impact of an education lottery on government finances (Section 2.3), donors' contributions (Section 2.4), and donations received by nonprofits (Section 2.5). The generic empirical specification employed throughout is:

$$y_{ist} = \alpha + edulot_{ist} + X_{ist} + [state\ FE's]_s + [year\ FE's]_t$$

where y_{ist} is the outcome variable of interest and X_{ist} is a vector of individual-level covariates. More importantly, "$edulot_{ist}$" is an indicator variable equal to one if observation *i* is in a state (*s*) that, at that point in time (*t*), sponsors an education-funding lottery.[15] Throughout, standard errors are adjusted to allow for clustering at the state-level. As noted in the previous section, the identification of an effect of introducing an education lottery stems from changes that occurred within the twelve states listed in Table 2.1.

There is some unavoidable imprecision in defining when the education lottery "treatment" begins in some cases. It is unclear, for instance, whether we should expect a lottery introduced in November of 2005 to generate observable changes in donations or government spending during the entire 2005 calendar year. Thus, a lottery that is introduced in the second half of the year (after June 30) is coded as beginning in the following year. Similarly, data from nonprofit tax returns and government finances are reported by fiscal year rather than calendar year. Thus, data from states or nonprofits with fiscal years ending in the first half of the calendar year (prior to June 30) is interpreted as data from the preceding year, which is the year during which the majority of the relevant activity took place.

One potential concern with the difference-in-differences approach is that states may introduce education lotteries in response to a decrease in the availability of education funding from either private or public sources. This would violate the assumption of parallel pre-treatment trends across treatment and control states. However, factors that are unrelated to education financing (e.g., within-state religiosity, the adoption of a lottery in a neighboring state) have been shown to be more important predictors of lottery adoption than fiscal crises (Coughlin et al., 2006), particularly in lotteries introduced after the 1970s (Alm et al., 1993). Additionally, in the next section I offer evidence that lottery states do not experience drops in revenue or increases in education expenditures in the years preceding to the adoption of a lottery.

[15] An "education-funding lottery" is defined here as a lottery that is introduced solely for the purpose of funding education. Some states defined here as education lotteries use a small fraction of their revenues for other causes, but only after achieving a certain threshold of funding for education. Thus, more precisely, an education lottery is defined here as a lottery for which the entire first dollar of revenue is earmarked for education.

2.4 STATE LOTTERIES & GOVERNMENT FINANCES

In this section, I examine the impact of the introduction of an education-funding lottery on state expenditures and revenues. The main focus of the paper is the analysis of charitable donations (Sections 2.5 and 2.6). This section is included both as an introduction to the lottery "treatment" (as it is implemented by government) and to allow for considerations of the impact of a lottery on *overall* education funding, accounting for both charitable donations and government spending.

2.4.1 Existing literature on fungibility of earmarked lottery revenue

If education lotteries crowd out charitable donations to education without an accompanying increase in state expenditures, then the true impact of an education lottery is a reduction in overall education funding. As noted, a handful of studies have examined the extent to which earmarked lottery revenue increases government expenditures for the intended beneficiary. While all of the papers in this literature find at least some evidence of fungibility, the degree of fungibility varies widely across studies.

Novarro (2005) and Evans and Zhang (2007) find that, conditional on sponsoring a lottery, an additional dollar of lottery proceeds intended for education does increase K-12 education expenditure, but by significantly less than a dollar. For instance, Evans & Zhang estimate that a dollar of lottery proceeds generates $0.50-$0.70 of education spending. Garrett (2001), Spindler (1995) find that earmarked lottery spending is completely offset by decreases in spending from other revenue sources so that total spending does not change. Borg et al. (1991), Erekson et al. (2002) find that earmarking lottery revenues actually *decreases* total spending on education.

However, some of these studies either compare state expenditures for just one state (or a handful of states) before and after a lottery (Garrett, 2001; Spindler, 1995) or, alternatively, for all 50 states but within just one year in a cross-section (Borg et al., 1991; Erekson et al., 2002). It is difficult to determine whether the effect (or lack of effect) of the lottery is causal. Novarro (2005) and Evans & Zhang (2007) are exceptions to this as they construct panels of all fifty states and employ state fixed effects; however, both focus is on dollar-for-dollar changes in spending conditional on sponsoring a lottery rather than the total average change in spending after the introduction of a lottery. Some of Novarro's results at least seem to suggest that there is little *overall* change in spending. More importantly, both Novarro (2005) and Evans & Zhang (2007) study a different outcome variable (elementary education spending) and an earlier time period (1976-2000) than the present research. Thus, before proceeding to the analysis of charitable contributions, I begin by assessing the impact of education lotteries on government finances in a panel of all 50 states during the time period studied in the remainder of the paper (1989-2008).

2.4.2 Data and empirical approach

I use data from the Census Bureau's *Survey of Government Finances*. The *Survey* provides a yearly account of states' expenditures and revenues, broken into detailed categories. From this data, I construct a panel of all fifty states from 1989 through 2008 where each observation represents a particular state-year combination. I follow the difference-in-differences framework discussed in Section 2.3 and estimate fixed-effects regressions, with fixed-effects at the state level. The dummy variable "Edulot," indicating that the state operates an education-funding lottery, is of primary interest. I examine the impact of being treated on several revenue and expenditure measures; all outcome variables are in logs and measured at a per capita level. All specifications in this subsection include controls for available time-varying state-level variables that may impact state revenues and expenditures: log of population, log of per capita income, a dummy for the presence of a non-education lottery, and (when revenue is not an outcome variable) log of non-lottery revenue.

2.4.3 Results

The impact of an education lottery is reported in Table 2.2, with each column taking a different measure of revenue (columns 1-3) or expenditure (columns 4 & 5) as the dependent variable. Education lotteries are successful in increasing state revenue: overall revenue increases by roughly 3% with the introduction of a state lottery (column 1); non-tax revenue (which still includes lottery revenue) increases by 7% (column 2). One might be concerned that states introduce new taxes in the same year that they introduce a lottery as part of a broader funding initiative; this would be problematic for our interpretation of the cause of crowded-out charitable contributions in future sections. However, we see in column 3 that this is not the case.

We now turn to the impact of lotteries on expenditure. Most importantly, we see that the introduction of an education lottery does not significantly increase education spending (column 4). The introduction of an education lottery *is* associated with an increase in non-education related spending (column 5). Non-education spending is defined here as total spending minus education related spending. The model predicts that non-education spending increases by 5.6%.[16]

Table 2.2. DID Estimates of state revenue and expenditure response to education lottery (FE-Reg.)

VARIABLES	(1) Revenue	(2) Non-tax revenue	(3) Tax revenue	(4) Education expenditure	(5) Non-education expenditure
Edulot	0.0302*	0.0692**	-0.0235	-0.00932	0.0561**
	(0.0159)	(0.0265)	(0.0251)	(0.0312)	(0.0230)
Population	-0.402***	-0.426***	-0.359***	-0.214**	-0.179
	(0.0714)	(0.156)	(0.0985)	(0.106)	(0.191)
Income per capita	0.975***	0.657***	1.514***	0.581**	0.493***
	(0.192)	(0.241)	(0.304)	(0.226)	(0.166)
Non-educ. lot.	-0.00509	0.0210	-0.0393	-0.0460	0.0292
	(0.0236)	(0.0325)	(0.0371)	(0.0412)	(0.0259)
Other revenue				0.224***	0.258***
				(0.0759)	(0.0598)
Constant	4.174***	11.71***	1.236	1.077	1.449
	(1.311)	(2.489)	(1.774)	(1.744)	(2.875)
State FE's	X	X	X	X	X
Year FE's	X	X	X	X	X
Observations	950	950	950	950	950
R-squared	0.962	0.936	0.934	0.953	0.978

Robust standard errors (clustered at the state-level) in parentheses.
All continous controls and outcome variables are in logs and measured at the per-capita level.
*** p<0.01, ** p<0.05, * p<0.1

[16] The main results from Table 2.2 are robust to the inclusion of additional time-varying controls and the inclusion of pre-treatment and post-treatment trends. See Appendix Table A1.

Relative to the existing literature, these empirical specifications are closest to those of Novarro (2005) and Evans & Zhang (2007), in the sense that they assess the impact of earmarking in a panel with state fixed effects. Both of their papers find that, while there is evidence of fungibility, there is *some* positive relationship between lottery revenue and education spending. Thus, while charitable donations are the main focus of this paper, the fact that I observe no increase in government education expenditure warrants a brief comment. First, both of their papers take elementary (K-12) education expenditures as the dependent variable and focus on an earlier time period (1976-2000). If I restrict my analysis to elementary education spending and shift my analysis to the years they cover (reported in Appendix Table A2), I too find a positive impact of an education lottery. However, this still fails to generate a significant increase in overall education expenditures. Thus, the extreme fungibility observed here may be a more recent phenomenon.

Table 2.3. State expenditure response to education lottery **--** other expenditure categories (FE-Reg.)

Expenditure category	Mean per capita spending (2008 dollars)	Treatment effect (Edulot)	Expenditure category	Mean per capita spending (2008 dollars)	Treatment effect (Edulot)
Elem. educ.	*$905.76*	-0.0574	Public safety	*$187.76*	-0.00468
	(302.33)	(0.0578)		*(73.81)*	(0.0560)
Higher educ.	*$607.94*	0.0208	Health	*$177.23*	-0.0312
	(168.52)	(0.0355)		*(86.14)*	(0.0695)
Public welfare	*$1077.62*	0.0715	Hospitals	*$152.37*	0.0741
	(387.78)	(0.0508)		*(97.84)*	(0.158)
Govt. salaries	*$864.53*	0.0122	Interest on debt	*$174.51*	0.0661
	(451.90)	(0.0402)		*(142.83)*	(0.100)
Insurance trust	*$527.19*	0.130**	Utilities	*$38.12*	-0.380
	(245.12)	(0.0626)		*(93.71)*	(0.227)
Highways	*$418.10*	-0.00516	Financial admin.	*$83.73*	0.227*
	(206.58)	(0.0452)		*(50.00)*	(0.120)

What non-education expenditure areas *are* benefitting from lottery revenue? In Table 2.3, the empirical specification employed in Model 4 of Table 2.2 is repeated but with the top ten (non-education) expenditure categories as dependent variables. I also report the predicted impact of the lottery on higher and K-12 education expenditures separately. The introduction of an education lottery has the most impact on insurance trust expenditures[17] and financial administration[18], both of which significantly increase by more than 10%.

In Table 2.4, I assess the parallel trends assumption necessary for the difference-in-differences approach and also examine changes in revenue and education expenditure over time more generally. I regress expenditures and revenues on a dummy indicating that an education lottery is in place ("Edulot"), but I also include a dummy set to one 2 years prior to the introduction of a lottery ("Edulot – 2 years") and a dummy set to one 2 years after a lottery ("Edulot + 2 years"). Included (but not displayed) are the same controls included in preceding specifications (population, income per capita, presence of a non-education lottery, state fixed effects, year fixed effects, and – in Models 3-5 only – non-lottery revenue.)

The variable "Edulot – 2 years" is included to test whether there is a change in government revenue or education expenditures in the years leading up to the adoption of a lottery, which – as noted in the previous section – might impact our interpretation of the "treatment effect" associated with the introduction of a lottery, both here and in the remainder of the paper. However, we see that there is no significant difference between soon-to-be-treated and untreated states with respect to revenue (Models 1-2), education expenditures (Model 3), or either of the non-education expenditure categories that appear to benefit most from education lotteries (Models 4-5).

It is possible that it takes time for lottery revenues to be funneled into education spending. If this were the explanation for the absence of an increase in education expenditures, we would expect a significant and positive coefficient on "Edulot + 2 years." This is not observed (Model 3).

Of course, a related concern – particularly given the focus on donations in this paper – is that a lottery is introduced because private support for education (donations) is falling. However, using an empirical specification and data that will be discussed in more detail in Section 2.6, this

---

[17] Defined by the Census Bureau as "Cash payments to beneficiaries (including withdrawals of retirement contributions) of employee retirement, unemployment compensation, workers' compensation, and disability benefit social insurance programs." Though not displayed here, I can demonstrate that this result is not entirely driven by employees' retirements – which would include teacher pensions.

[18] Defined by the Census Bureau as "Activities involving finance and taxation. Includes central agencies for accounting, auditing, and budgeting; the supervision of local government finances; tax administration; collection, custody, and disbursement of funds; administration of employee- retirement systems; debt and investment administration; and the like."

does not appear to be the case (Appendix Table A3). There are no observable differences in donations to education across treated and untreated states two years prior to treatment.

Table 2.4. Robustness test **-** State revenue and expenditure response to education lotteries in years preceding and following treatment

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Revenue	Non-tax revenue	Education expenditure	Insurance trust expenditure	Financial admin. expenditure
Edulot – 2 years	-0.00811	-0.00293	-0.0139	0.0136	0.0108
	(0.0268)	(0.0424)	(0.0195)	(0.0388)	(0.0892)
Edulot	0.0245	0.0519*	-0.00265	0.0714	0.251***
	(0.0175)	(0.0282)	(0.0170)	(0.0484)	(0.0645)
Edulot + 2 years	0.0198	0.0331	0.00623	0.0825*	-0.0546
	(0.0176)	(0.0206)	(0.0154)	(0.0481)	(0.0831)
Observations	950	950	950	950	950
R-squared	0.962	0.937	0.953	0.931	0.881

Robust standard errors (clustered at the state-level) in parentheses.
All outcome variables are in logs and measured at the per-capita level.
Specifications include additional controls included (but not displayed) as noted in the text.
*** p<0.01, ** p<0.05, * p<0.1

To summarize, education lotteries significantly increase revenue but fail to significantly increase education expenditures for education lotteries introduced between 1989 and 2008. This is consistent with most of the existing literature on the fungibility of earmarked lottery revenues. Thus, overall funding available for education hinges on the impact of an education lottery on private donations to education-related causes, which we examine next.

2.5 DONOR RESPONSE TO LOTTERIES

2.5.1 Data and empirical approach

How do donors respond to the introduction of an education lottery? To begin to answer this question, I analyze responses from three individual-level surveys: the Giving and Volunteering in the United States Survey (GVS), the Center on Philanthropy Panel Study (COPPS), and the

Consumer Expenditure Survey (CES). All three of these surveys ask respondents to indicate how much they have donated to a variety of causes, including education.

GVS and COPPS were designed to gather information about individuals' charitable activities and are two of the most widely used sources of data on the topic. Both surveys ask detailed questions about the amount donated to various charitable causes such as education, health, public services, etc., in addition to more basic demographic information. COPPS follows a panel of individuals between 2001 and 2009 (with surveys every two years). GVS is not a panel, but I have constructed a repeated cross-section of surveys between 1990 and 1999 (again, with waves every two years – until 1996, when the next wave was not administered until 1999). In both surveys, participants are asked about their charitable giving in the preceding calendar year, so collectively GVS and COPPS provide results for the years 1989 through 2008.

The COPPS data is preferable as it is a panel and allows for individual fixed effects, thereby controlling for unobserved differences in altruism. However, given that identification in the difference-in-differences framework stems from a state establishing an education lottery within 2001 to 2009, one might be concerned that the results are driven by something specific about this handful of states. Thus, the GVS data is included to further support the robustness of the results by providing additional observations during a different decade with different states introducing education lotteries.

The Consumer Expenditure Survey (CES) – which follows respondents for four quarters and rotates in a new wave of respondents each quarter – is of course not primarily designed to address charitable giving. However, the survey does ask respondents to indicate how much they give to charitable organizations (which they define as organizations "such as United Way, Red Cross, etc."), religious organizations, political organizations, and education organizations. Moreover, since 2001 the survey has asked respondents to indicate how much they spend on "Lotteries and games of chance." Neither GVS nor COPPS ask participants to indicate their lottery expenditures. Thus, CES (from 2001-2008) is used here to examine how the introduction of a lottery differentially impacts individuals who do and do not play the lottery. Because CES is a quarterly dataset, respondents are considered "treated" if there is an education lottery in their state in the *quarter* (as opposed to year) of response.

CES indicates respondents' state of residence. However, a drawback of the CES is that if there are too few respondents from a particular state, then it is not possible to identify state of residence for anyone from that state. Thus, while six states introduce an education lottery during the time period covered by CES, there are only observations from four of these states. Because of this, CES is primarily used to test the relationship between charitable expenditures and lottery expenditures.

The primary outcome variable of interest in all three datasets is total giving to education. In both COPPS and GVS, respondents report education giving for the preceding calendar year. In CES, respondents report education giving for the preceding quarter. In all three datasets, a handful of exceedingly high donations are removed; given the relatively low number of individuals who make positive contributions, a few contributions may wield excessive influence over the estimated mean impact of treatment.[19] To apply a consistent rule across all three datasets, observations with contributions above the 99th percentile of education contributions (conditional on making a positive education) contribution are removed. This amounts to 55 of the total 221,067 observations in CES, 52 of the 39,795 observations in COPPS, and 15 of the 12,133 observations in GVS. This allows me to estimate the impact of an education lottery on the "typical" donor; Section 2.6 – where I analyze data from nonprofit tax returns – provides the opportunity to assess the impact of the lottery on contributions at a more aggregate level.

Included covariates vary by the survey and empirical approach being used. In the COPPS data, all specifications control for family income. When the COPPS data is estimated without individual fixed-effects, I also include a variety of additional controls: number of children, employment (respondent and spouse), marital status, urban-rural residence status, age, sex, and race. GVS specifications include controls for race, gender, employment (respondent and spouse), church attendance, age, education level, income, marital status, children in household, and confidence in education (as indicated in the survey). Finally, in analyzing the CES data, I control for education level, total consumption expenditures, education level, age, sex, and race. Regardless of the survey being used, all specifications include year fixed effects and state fixed effects (unless the specification includes individual-level fixed effects.)

[19] For instance, in the GVS data, the inclusion of these extreme observations leads to estimates that suggest that the average drop in giving associated with an education lottery is larger than the initial mean of giving.

2.5.2 Results

Results from the baseline specifications in both datasets are presented in Table 2.5. Column 1 reports the results of a fixed-effects regression in the COPPS data; column 2 reports the results of the repeated cross-section analysis in the GVS data. In either case, we find that education giving significantly decreases when an education lottery is introduced. To provide some sense of the magnitude of these coefficients, the mean of education giving is roughly $40 in both datasets.

Table 2.5. Baseline results - Impact of education lottery on education giving

VARIABLES	(1) Educ. giving	(2) Educ. giving	(3) Any educ. giving	(4) Any educ. giving
Edulot	-9.372*	-33.49***	-0.033	-0.003
	(5.153)	(11.94)	(0.032)	(0.019)
Observations	29,715	11,017	7,985	11,012
Dataset	COPPS (2000-2008)	GVS (1989-1998)	COPPS (2000-2008)	GVS (1989-1998)
Model	FE Reg.	OLS	FE Logit	Logit

Robust standard errors (clustered at state-level) in parentheses
Columns 3 and 4 report marginal effects.
*** p<0.01, ** p<0.05, * p<0.1

I also estimate logit models to assess how an education lottery impacts giving on the extensive margin. Results are presented in columns 3 and 4 for COPPS and GVS, respectively. In both models, the dependent variable is equal to one if the respondent reports any education giving. There is little response to the introduction of an education lottery. Thus, the baseline results are driven by changes on the intensive margin. This provides an initial indication that these results may not be entirely driven "fundraiser crowd-out." If the only reason that contributions decrease is a decline in the number of donors being solicited, then we might expect to find that the drop in giving is driven by the extensive margin.

There is reason to be concerned that the simple baseline results might be biased due to the large number of individuals who contribute nothing to education. This concern is addressed in two ways, with results reported in Table 2.6. In analyzing the COPPS data, I can restrict the sample to "education-givers" – individuals who donate to education at *any point* in the panel.

This substantially reduces the number of zero-contribution observations. This specification is also interesting in its own right as it estimates the impact of the treatment on the individuals who *would* be giving. As the GVS is not a panel, the GVS parallel to this is to restrict the sample to observations with positive education contributions.[20] Results from these estimations are reported in Columns 1 and 3 respectively. Again, we see a significant decrease in giving in both datasets/decades but, as we would expect, the magnitude is much larger than the baseline result.

Table 2.6. Alternative specifications **--** Impact of education lottery on education giving

VARIABLES	(1) Educ. giving (Educ. givers only)	(2) Educ. giving	(3) Educ. giving (Educ. givers only)	(4) Educ. giving
Edulot	-27.77*	-8.44*	-191.9***	-11.96*
	(16.27)	(5.12)	(53.09)	(7.01)
Observations	9,279	28,426	1,801	11,017
Dataset	COPPS	COPPS	GVS	GVS
Model	FE	Tobit	OLS	Tobit

Robust standard errors in parentheses (clustered at state level in Models 1 and 3, individual level in Models 2 and 4). Columns 2 and 4 report marginal effect on the unconditional expected value of observed giving.
*** p<0.01, ** p<0.05, * p<0.1

In Models 2 and 4, I estimate Tobit models to address "censoring" of contributions at $0. There is not a straightforward and unbiased implementation of fixed effects in Tobit models for panel data, so in the COPPS data I instead estimate a standard Tobit model, adjusting standard errors for clustering at the individual-level (Column 3)[21]. Similarly, in Column 4 I report the results of estimating a Tobit model in the GVS data. For each of these specifications, I report the marginal effect of "Edulot" on the unconditional expected value of *observed* giving. In both cases, we continue to observe a significant decrease in giving after accounting for the large number of censored observations. These estimates suggest that the introduction of an education lottery decreases average giving by between eight and twelve dollars; from an average of $40, this represents a drop in giving of between 20 and 30 percent.

[20] Restricting our attention to education givers would be problematic if the treatment changed *the set of donors* and not just the size of their contribution.
[21] Estimating a random-effects Tobit model yields similar results.

What is driving this drop in giving? The decrease is consistent with classic models of crowd-out; the expected introduction of a new source of funding for a public good serves as a substitute for individual contributions and as such donors reduce their level of giving. There are of course alternative explanations. Kearney (2005) finds that, for the average lottery player, lottery spending is entirely financed by a reduction in non-gambling expenditures; thus, it is reasonable to expect that lottery spending may come at the expense of a particular category of non-gambling expenditures: charitable giving. If this were the case, we would expect charitable giving to decrease generally instead of finding a drop only in education-related giving.

Table 2.7. Giving to other non-education related causes

VARIABLES	(1) Non-educ. giving	(2) Non-educ. giving	(3) Non-educ. giving	(4) Non-educ. giving
Edulot	68.98	35.61	-25.38	22.69
	(73.38)	(47.70)	(47.16)	(32.15)
Observations	29,715	28,426	11017	11,017
Dataset	COPPS	COPPS	GVS	GVS
Model	FE-Reg	Tobit	OLS	Tobit

Robust standard errors in parentheses (clustered at state level in Models 1 and 3, individual level in Models 2 and 4). Columns 2 and 4 report marginal effect on the unconditional expected value of observed giving.
*** p<0.01, ** p<0.05, * p<0.1

It is not the case that giving to other causes substantially decreases with the introduction of an education lottery. To show this, I again estimate the baseline specifications (Columns 1 and 2 of Table 2.5) and the Tobit models (as in Table 2.6) but take "non-education giving" as the dependent variable.[22] Results are reported in Table 2.7. With the exception of Column 3 (where there is a small and insignificant drop in non-education giving), we see that giving to other causes actually slightly increases, but this increase is not significant. (The magnitudes of these coefficients are larger than those of the education-only estimations as the mean of giving to the

[22] In GVS, "non-education giving" is defined as total reported giving minus education giving. In COPPS, respondents do not report "total giving" and the way that they are asked to report giving to several causes changed between the 2002 wave and the remaining waves. However, questions regarding education giving, religious giving, "combined purpose" giving (e.g., United Way), health giving, and "help for the needy" are consistent across waves. Thus, in COPPS "non-education giving" is the sum of these consistently measured categories (religious, health, combined purpose, and needy).

sum of other causes is naturally much higher than giving to just education. For instance, average non-education giving in GVS is roughly $224.)

Another alternative explanation is that donors view the lottery as a good substitute for education donations. This is the assumption made by Morgan (2000) in claiming that fixed prize raffles and lotteries can lead to higher provision than voluntary contributions; in his model, one dollar of lottery spending increases public good provision by the same amount that one dollar of direct donation does. However, lottery "contributions" are subsidized, in that there is some chance of winning a prize. As a result, his model predicts that individuals who *would* be donating to education in the absence of a lottery will shift their donation expenditures to the lottery. Because of this, *direct* donations drop but *total* contributions (donations + lottery expenditures) actually increase. Thus, a decrease in direct donations – as has been demonstrated thus far – arguably does not sufficiently demonstrate that overall public support for education has fallen.

If a substitution from donations to lottery expenditure was driving the results, then we would expect the decrease in giving to stem entirely from individuals who play the lottery. Using the Consumer Expenditure Survey (CES) we can identify individuals who do and do not play the lottery, and how much each group contributes to education related causes. Recall that the CES provides quarterly observations of respondents throughout the course of a year. Thus, I identify an individual as a "lottery nonparticipant" if they never report positive lottery expenditures. A "lottery participant" reports lottery expenditures at some point during the year.

Using these data, I estimate fixed-effects regressions (Column 1 of Table 2.8) and (cross-sectional) Tobits (Column 2), taking *quarterly* education giving as the dependent variable. I also estimate fixed-effects logits to assess the likelihood of giving any positive amount to education (Column 3). In the fixed-effects regression and logit, I include a control for income and also quarter, state, and year fixed effects. In addition to these controls, the Tobit also includes controls for education, gender, and race. I conduct each estimation for three samples: all observations (Panel A), lottery nonparticipants (Panel B), and lottery participants (Panel C).

Comparing Panels B and C, there is more evidence that the drop in giving is driven by lottery *nonparticipants* in this setting. The estimated impact of the lottery in the fixed effects model is negative for both participants and nonparticipants, but the magnitude is larger (and more precisely estimated) for nonparticipants. Based on the Tobit estimations, the impact of the lottery remains negative and significant for nonparticipants, but is positive and insignificant for participants. Finally, although the logit estimates are insignificant for all samples, the impact of the lottery on the extensive margin is again greater for nonparticipants. The generally larger negative impact of the lottery for *nonparticipants* and weak impact for participants is the opposite of what we expect if donors simply perceived the lottery as an alternative way to contribute. Moreover, these results paired with the results from Table 2.6 provide additional evidence that the drop in giving is not simply a more general shift of expenditures from charity to lottery. This suggests that the main result is in line with classic models of crowd-out; donations drop because there is a new source of revenue (lottery revenue) that serves as a substitute for one's own donations.

Table 2.8. Consumer Expenditure Survey **--** Quarterly education giving by lottery participation

	VARIABLES	(1) Educ. giving	(2) Educ. giving	(3) Any educ. giving
	Model:	FE-Reg.	Tobit	FE-Logit
Panel A:	Edulot	-22.18**	-0.103	-0.077
Full sample		(8.593)	(1.75)	(0.081)
	Observations	203,486	203,510	12,624
Panel B:	Edulot	-24.30*	-0.828***	-0.133
Lottery nonparticipant		(13.80)	(0.944)	(0.091)
	Observations	153,087	153,099	8,488
Panel C:	Edulot	-16.92	0.636	0.039
Lottery participant		(10.45)	(2.64)	(0.175)
	Observations	50,399	50,411	4,136

Robust standard errors in parentheses
Columns 2 and 3 report marginal effects
*** p<0.01, ** p<0.05, * p<0.1

To summarize, the introduction of an education lottery reduces donors' contributions to education lotteries by 20 to 30 percent. This reduction appears to be driven by changes on the intensive margin; the lottery does not impact the probability that an individual will make a contribution. Moreover, there is evidence that the drop in giving might be explained by (expected) government spending crowding out private contributions, as opposed to individuals sacrificing charitable contributions to play the lottery. However, while the decrease in giving seems to be a response to new government funds, it remains unclear whether this is a response by *donors* or a response by *nonprofit firms*. This issue is explored in the next section.

2.6 NONPROFIT FIRM RESPONSE TO EDUCATION LOTTERY REVENUE

How does the introduction of an education-funding lottery impact education-related nonprofits? We have already seen that an education lottery crowds out donations to education organizations, but it is possible that the result is driven by a reduction in the effort of fundraisers – either because they expect the marginal benefit of fundraising to be lower or because they have benefitted directly from lottery revenues and their level of need has reduced. Andreoni and Payne (2003) document that, in a more general setting, the crowd-out that results from government grants to nonprofits can almost be entirely explained by this "fundraising crowd-out." In some of their results that account for fundraising, donors' contributions are either unaffected by or slightly increase with grants. One explanation for their observed lack of "traditional crowd out" (and, more generally, for the relatively small degree of crowd-out typically observed in response to government grants) is that individuals are largely unaware of government grants to nonprofits. In the United States, the introduction of a lottery to fund education tends to be highly publicized and as such individuals are more likely to be aware of this change in government funding. Thus, it may be reasonable to expect that the crowd-out observed in the previous section is driven by donor preferences.

2.6.1 Data & empirical approach

To examine whether this is the case, I next turn to data on nonprofit organizations' revenue and expenses from federal tax returns spanning from 1989 to 2007. The data is collected and constructed by the IRS Statistics of Income division, and then compiled and provided for research purposes by the National Center for Charitable Statistics (NCCS). Each year a subset of tax returns from nonprofit organizations that hold 501(c)(3) status are randomly sampled for inclusion in the dataset, which reports a variety of financial variables from the their tax return (from the year sampled) such as operation expenses, charitable contributions received, fundraising expenses, etc. The dataset also includes groupings of nonprofit organizations by function, categorizing organizations as *Arts, Education, Health, Human Services,* or *Other*.

A broad array of education-related organizations are represented in the data, including colleges, universities, preschools, libraries, remedial reading organizations, etc. However, as noted in the introduction, a vast majority of charitable activity in the education subsector is directed towards higher education. These data of course include private nonprofit colleges and universities, but many public universities and colleges are also represented: either because (1) they officially hold 501(c)(3) status or (2) their fundraising activities are accomplished through an affiliated but independent nonprofit foundation, both of which are common.

While the dataset is not constructed as a true panel of nonprofit organizations, nonprofit organizations reappear in the data often enough that it can be treated as panel (as Andreoni & Payne do, for instance).[23] Thus, I construct an unbalanced panel where each observation is a particular nonprofit firm in a particular year; there are typically (but not always) gaps between a nonprofit firm's appearances in the panel but these appearances are randomly determined.

The goal of this section is to examine the donations received and the fundraising behavior of nonprofit firms in response to the introduction of a lottery. Thus, I restrict my sample to firms that receive donations at any point in the panel. The resulting dataset consists of a total of 192,478 observations and 19,505 unique nonprofit firms. 39,410 of these observations are education-related organizations.

Throughout this section, I use a fixed-effects approach (with fixed-effects at the firm level) within the same difference-in-differences framework employed in previous sections. Two questions are of primary interest: First, how does the introduction of an education lottery impact the amount of donations received by education-related organizations? This essentially tests the robustness of the results from the previous section, but with much richer data. Here, for instance, we do not suffer from the censoring at $0 that plagued the assessment of the donor-level data. Second, are changes in donations received by nonprofits driven by changes in fundraising efforts?

[23] The median firm in the dataset I use appears seven times.

With these questions in mind, the primary outcome variables of interest is log of *contributions received*. I regress contributions on the "Edulot" indicator variable and, in all specifications, I include controls for the log of total revenue (excluding public support), the log of total expenditures (excluding fundraising), and year fixed effects. I additionally control for state-level covariates which may impact donations: log of income per capita, log of state population, log of education expenditure per capita, and log of other expenditures per capita. To address the impact of fundraising, I then control for fundraising expenditures. In doing so, I use an instrumental variables approach to account for the endogeneity between fundraising and donations received, using *liabilities at the beginning of the fiscal year* as an instrument for fundraising.[24]

2.6.2 Results

Table 2.9 provides an initial assessment of the impact that a lottery has on contributions received by nonprofits. Columns 1 and 2 report the results of fixed-effects estimations, with fixed-effects at the firm level, for education organizations and non-education organizations respectively. Consistent with the findings from the previous section, the introduction of an education lottery reduces the contributions received by education organizations – in this case, by an estimated 8% – but has no significant impact on contributions to other causes.

Columns 3 and 4 offer two robustness tests. Between 1989 and 2008, three states[25] attempted to introduce a lottery through referenda or ballot initiatives, butto achieve enough votes. In Column 3, I replace the "Edulot" dummy with a "Failed edulot" dummy. If the treatment effects here are merely picking up trends in giving that *cause* a state to introduce a lottery, the coefficient on "Failed edulot" should be negative and significant. While negative, the coefficient is substantially smaller in magnitude than the result from Column 1 and is not significantly different than zero. Column 4 adds a dummy to indicate a non-education lottery to

[24] Andreoni & Payne (2011) use this instrument for fundraising as well arguing that higher debt impacts the need for fundraising in a way that is unrelated to the amount of donations one expects to receive.
[25] Oklahoma in 1995, Alabama in 2000, Arkansas in 2001.

the main specification. No drop in giving to education organizations is observed when the lottery is not intended to benefit education.

Table 2.9. The impact of an education lottery on contributions received

VARIABLES	(1) Contributions received: **Educ. orgs.**	(2) Contributions received: **Non-educ. orgs.**	(3) Contributions received: **Educ. orgs.**	(4) Contributions received: **Educ. orgs.**
	FE-Reg.	FE-Reg.	FE-Reg.	FE-Reg.
Edulot	-0.0817***	0.0119		-0.0734*
	(0.0275)	(0.0282)		(0.0369)
Failed Edulot			-0.0247	
			(0.0817)	
Non-educ. lottery				0.0138
				(0.0363)
Other expenditures	0.339***	0.282***	0.340***	0.339***
	(0.0374)	(0.0142)	(0.0373)	(0.0374)
Other revenues	-0.0747***	-0.0659***	-0.0745***	-0.0747***
	(0.0216)	(0.00876)	(0.0217)	(0.0217)
State: income	0.850***	1.053***	0.783***	0.851***
	(0.258)	(0.294)	(0.258)	(0.258)
State: population	0.433	0.666***	0.264	0.425
	(0.270)	(0.198)	(0.241)	(0.280)
State: educ. exp.	0.0443	0.0149	0.0565	0.0435
	(0.105)	(0.0889)	(0.108)	(0.104)
State: non-educ. exp.	-0.131	-0.217*	-0.149	-0.135
	(0.140)	(0.115)	(0.145)	(0.139)
Observations	38,585	129,267	38,585	38,585
R-squared	0.263	0.066	0.227	0.331

Robust standard errors (clustered at state-level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Is the decrease in contributions to education organizations driven by a change in fundraising efforts? To answer this, I add a control for fundraising to the preceding specification. However, to account for potential endogeneity between fundraising and donations, I do so in an instrumental variables framework, taking liabilities as an instrument for fundraising. Results for education (Columns 1 and 2) and non-education organizations (Columns 3 and 4) are presented in Table 2.10. Columns 1 and 3 report the first stage of the instrumental variables regression. Notably, the introduction of an education lottery has very little impact on education organizations' fundraising expenditures (Column 1). Thus, in turning to the impact of the lottery after accounting for fundraising (Column 2), it is unsurprising to find that the estimated decrease in giving is very close to the estimate from Table 2.9.

Table 2.10. The impact of an education lottery accounting for fundraising

	(1) Educ. orgs: Fundraising	(2) Educ. orgs: Contributions received	(3) Educ. orgs: Contributions received	(4) Non-educ. orgs: Fundraising	(5) Non-educ. orgs: Contributions received	(6) Non-educ. orgs: Contributions received
	FE-Reg. (first-stage)	IV-FE-Reg.	FE-Reg.	FE-Reg. (first-stage)	IV-FE-Reg.	FE-Reg.
Edulot	-0.00296	-0.0689**	-0.0650**	0.0685	0.00314	0.0247
	(0.0324)	(0.0279)	(0.0296)	(0.0422)	(0.0412)	(0.0372)
Liabilities	0.0346**			0.0432***		
	(0.0130)			(0.00644)		
Fundraising		0.0614	0.130***		0.4851***	0.176***
		(0.2656)	(0.0116)		(0.1273)	(0.00924)
Obs.	27,905	27,905	28,828	58,374	58,374	61,996
R-squared	0.452	0.155	0.512	0.245	0.469	0.363

Robust standard errors (clustered at state-level) in parentheses
Additional controls included as noted in text
*** p<0.01, ** p<0.05, * p<0.1

On average the introduction of an education lottery reduces contributions received by education organizations by between 7-8%. To link this result more closely to the existing literature on crowding-out of charitable giving, we would ideally like to know the extent to which charitable giving decreases as a function of the *amount* that government spends. Answering this question is difficult because there is very little actual increase in spending, but we do know how much government *claims* it will spend. That is, in the state government finance data I observe "lottery proceeds," which is the amount of money remaining for the beneficiaries after accounting for prizes awarded and administrative costs. Thus, we can estimate the continuous impact of treatment by adopting the same specifications as before but replacing the "Edulot" dummy with log of lottery proceeds in education-lottery states. Appendix Table A4 reports the results of these estimations for both education and non-education organizations. Based on the instrumental variable specification which controls for fundraising (Table A4, Panel

B), a 10% increase in lottery proceeds is associated with a 5.25% decrease in contributions received by education related organizations.

2.6.3 Salience of government activity as an explanation for crowd-out?

Consistent with the findings from Section 2.5, contributions to education-related organizations fall after the introduction of an education lottery, which is not true of contributions to other organizations. However, we can now say that this result appears to be driven by donors' decisions to reduce their contributions as opposed to reduced fundraising efforts.

This result differs from a recent literature which demonstrates that crowd-out is often largely explained by a change in nonprofits' fundraising behavior (Andreoni & Payne, 2003; Andreoni & Payne, 2011; Heutel, 2009; Monti, 2010). I have suggested that an important difference between state lotteries and other forms of government spending is the high level of publicity that lotteries receive. Relative to government grants to nonprofits, donors are likely to be more aware of government spending resulting from lotteries – and therefore more likely to respond – in large part because states themselves heavily advertise the recipient of lottery revenues.

Is there more direct evidence to support this suggestion? I take two approaches answer this question. First, if the crowd-out observed in this paper is indeed driven by donors' awareness of government activity and if this awareness is (at least in part) the result of government advertising, then we would expect the magnitude of the crowd-out to increase with governments' advertising activities. The *Survey of Government Finances* (used in Section 2.3) reports states' yearly lottery administrative costs, which includes advertising expenditures.[26] Advertising expenditures are not reported, so I use the ratio of *administrative costs to ticket sales* as a proxy for advertising. In addition to advertising, administrative costs include the cost of printing and distributing tickets which obviously varies with the number of tickets sold, so most of the

[26] According to the Census Bureau, administrative costs "includes salaries of officials as well as advertising, supplies, and the like."

variation in administrative costs after accounting for tickets sales presumably comes from advertising.

I extend the previous empirical specifications (FE and FE-IV-Regressions controlling for fundraising) to include controls for *the ratio of administrative costs to ticket sales* ("Advertising") and the interaction of "Advertising" with "Edulot." In doing so, I re-center "Advertising" around its mean so that the main effect of "Edulot" can be interpreted as the impact of an education lottery evaluated at the mean level of advertising. If crowd-out is increasing in advertising we would expect the coefficient on "Edulot X Advertising" to be negative.

This is indeed the case, as can be seen in Columns 1 and 3 of Table 2.11 which report the results of these estimations for education organizations. Based on Column 3, an education lottery is associated with a 6% decrease in contributions received by education organizations. For each additional cent of ticket sales that a state devotes to administrative costs, contributions decrease by an additional 1%. The same significant relationship does not hold for non-education organizations (Columns 2 and 4).

Table 2.11. Crowd-out and awareness of government spending **-** Proxy for advertising expenditures

VARIABLES	(1) Contributions received: **Educ. orgs.** (FE-Reg.)	(2) Contributions received: **Non-educ. orgs**. (FE-Reg.)	(3) Contributions received: **Educ. orgs.** (FE-IV-Reg.)	(4) Contributions received: **Non-educ. orgs.** (FE-IV-Reg.)
Edulot	-0.0686**	0.0263	-0.0594**	0.0154
	(0.0286)	(0.0280)	(0.0256)	(0.0393)
Edulot X Advertising	-1.274***	-0.589	-0.988**	-0.247
	(0.387)	(0.485)	(0.466)	(0.488)
Advertising	0.655***	0.00285	0.556***	-0.243
	(0.200)	(0.270)	(0.197)	(0.221)
Observations	38,585	129,267	27,905	58,374
R-squared	0.294	0.068	0.156	0.000

Robust standard errors (clustered at state-level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

A second approach allows for the possibility that the political method of introducing the lottery impacts crowd-out. In particular, seven of the twelve states that introduced an education lottery between 1989 and 2008 did so through referenda or ballot initiatives.[27] The remaining states introduced their lottery through legislative action. One might expect that citizens are more aware of the lottery and its beneficiary when they vote directly on the issue. Thus, if salience is important to crowd-out, there should be more crowd-out in states that introduced their lotteries through direct voting (referenda/ballot initiatives).

I test whether this is the case in Table 2.12, which includes a separate treatment dummy for *legislative action* and *direct vote* states. Columns 1 and 3 report the results of these estimations for education organizations. Crowd-out is indeed higher in *Direct vote* states. The same relationship is not observed for non-education organizations (Columns 2 and 4).

Table 2.12. Crowd-out and awareness of government spending **--** Political method of lottery adoption

VARIABLES	(1) Contributions received: **Educ. orgs.** (FE-Reg.)	(2) Contributions received: **Non-educ. orgs**. (FE-Reg.)	(3) Contributions received: **Educ. orgs.** (FE-IV-Reg.)	(4) Contributions received: **Non-educ. orgs.** (FE-IV-Reg.)
Edulot (Legislative)	-0.0627*	-0.00842	-0.0413	-0.00215
	(0.0329)	(0.0296)	(0.0473)	(0.0468)
Edulot (Direct vote)	-0.0921**	0.0261	-0.0842***	0.00658
	(0.0355)	(0.0374)	(0.0275)	(0.0562)
Observations	38,585	129,267	27,905	58,374
R-squared	0.266	0.066	0.155	0.000

Robust standard errors (clustered at state-level) in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Of course, these results should be taken as merely suggestive: we cannot directly observe advertising expenditures, nor do we *know* that donors are more aware of the lottery beneficiary in "Direct voting" states. However, the results are consistent with the suggestion that a higher level of awareness of government activity leads to more crowd-out. This may help explain why crowd-out is driven by donors when the source of funding is a state lottery, while crowd-out is driven mostly by nonprofits when the source of funding is much-less-publicized government grants.

[27] These states are Georgia, Missouri, Virginia, Washington, South Carolina, Tennessee, and Oklahoma.

2.7 CONCLUSION

In this paper I assess the impact that education-funding state lotteries have on total funding available for education. I find that – for lotteries introduced between 1989 and 2008 – the introduction of an education lottery fails to significantly increase state education expenditures. Instead, unrelated expenditures increase. Thus, the lottery does not change government's contribution to education. Though education relies on government for funding, it is also heavily supported by charitable donations. The absence of an increase in government funding with the introduction of a lottery implies that any change in *overall* funding therefore depends on the effect that lotteries have on charitable contributions.

I find that charitable contributions to education significantly decrease after the introduction of an education lottery; contributions received by education-related nonprofit firms drop by 8% with a lottery. There is evidence to suggest that this drop is driven by a crowding-out of donations, consistent with classic models of voluntary public good provision. In particular, I am able to rule out alternative explanations that might suggest that individuals are merely shifting charitable expenditures to lottery expenditures.

Additionally, unlike recent work that finds that crowd-out stemming from grants to nonprofits is often mostly explained by nonprofit fundraising behavior, here the effect is almost entirely driven by donors. I argue that this is because of the high level of publicity that lotteries and their intended beneficiaries receive. Consistent with this suggestion, I show that crowd-out is increasing in a measure of state advertising activity. Also, crowd-out is higher for states that introduce a lottery through referenda instead of legislative action, which is presumably less salient to citizens. Though the potential importance of salience as a determinant of charitable crowd-out has been discussed in recent work by Monti (2010), to my knowledge this is the first paper to provide empirical evidence that crowd-out is indeed increasing in awareness of government activity.

There are of course a variety of policy-oriented reasons why some oppose state-sponsored lotteries; for instance, it has been repeatedly shown that, as a tax, lotteries are highly regressive. This paper highlights an additional trade-off that states face in implementing a lottery as a way to fund public goods. While some existing work shows that earmarking for a "good cause" increases a lottery's revenue (Landry & Price, 2007), I find that this comes at a price: private, voluntary support for the cause falls.

However, the fact that state governments are vocal about the *particular* cause being supported (education) seems to be critical to this result. This suggests that a government that is vocal about supporting "good causes," but does not support or highlight any one cause in particular, may enjoy the benefits of higher revenue without disrupting charitable activity. The UK National Lottery operates in this manner, advertising that the Lottery supports "380,000 … good causes … across the UK."[28] Indeed, in an analysis of UK charities that have received lottery grants, Andreoni et al. (2013) find no evidence of charitable crowd-out.[29]

---

[28] http://www.national-lottery.co.uk/player/p/goodcausesandwinners.ftl

[29] In fact, for small organizations, they find evidence of *crowd-in*.

3.0 THE SUPPLY & DEMAND OF MOTIVATED LABOR: WHEN SHOULD WE EXPECT TO SEE NONPROFIT WAGE GAPS?

3.1 INTRODUCTION

A long literature examines differences in wages across the for-profit and nonprofit sectors. Some empirical work finds clear evidence that nonprofit workers earn less. The most prominent explanation for this finding is the "labor donation hypothesis," which suggests that some individuals enjoy nonmonetary benefits from working in a nonprofit and as such are willing to work for less.[30] An alternative explanation is that nonprofits tend to locate in lower-paying industries, so composition effects rather than differential preferences of workers may drive the observation of a nonprofit wage differential. However, as demonstrated by Leete (2001) even after carefully accounting for industry and occupation, a nonprofit wage differential exists in some industries but not others – a puzzling result for either of these two prominent explanations.

I offer an explanation for the emergence of wage differentials in some industries but not others and, in doing so, revisit the mechanism through which these wage differentials arise more generally. Building on a point made by Preston (1989), I test the hypothesis that a nonprofit wage differential should exist within an industry when the share of labor demanded by nonprofits is low relative to for-profits. I start from the assumption that some workers are in fact willing to donate their labor and draw from a recent theoretical literature examining the impact of motivation on labor market outcomes (Besley & Ghatak, 2005; Heyes, 2005; Delfgaauw & Dur, 2007; Ghatak & Mueller, 2011). Suppose there exist "motivated" types – who receive

[30] "Labor donation theory" or the "donative labor hypothesis" is an idea which has been suggested in various forms by a variety of researchers; see, for instance, Weisbrod (1983), Preston (1989), or Leete (2006) for a thorough review.

nonmonetary benefits from working for a nonprofit – and "standard" types – who, holding wage constant, do not differentiate between nonprofit and for-profit jobs. As long as there are enough motivated workers to meet their labor demands, nonprofits can minimize costs by offering a low wage (thereby, only attracting motivated applicants.) However, if nonprofit labor demand is high relative to for-profit firms, the nonprofit cannot rely on motivated workers alone to fill their demand and must offer wages comparable to that of for-profits in order to attract standard workers.

I test this hypothesis empirically in two stages. In both stages, rather than attempting to make comparisons across very different industries, I examine the impact of the nonprofit share of labor *within* industries but *across* localities. The first stage assesses the relationship between nonprofit share of labor and wage differentials at an economy-wide level using data from the 2000 Public-Use Microdata 5%-sample of the United States Census. I construct industry/locality-specific nonprofit shares of labor and include a detailed set of industry fixed effects in all specifications. I find evidence that low nonprofit share – where I argue that there is a sufficient number of motivated workers to meet labor demand – is indeed associated with larger negative wage differentials; this is almost entirely driven by college-educated workers. However, there are a number of alternative explanations for this result that data limitations do not allow me to address. In particular, there are reasons to be concerned that there is endogeneity between nonprofit share and relative nonprofit wages. Additionally, in the economy-wide data, I cannot rule out that low nonprofit share nonprofits attract lower quality workers and the lower wage merely reflects this.

Thus, in the second stage, I focus my attention only on the nursing home industry, for which I have much richer firm-level data on roughly 95% of nursing homes in the United States. Again, exploiting variation in nonprofit share across localities, I replicate the result from the economy-wide data. I then provide evidence that suggests that this result is not driven by: (1) differences in the competitive environment faced by nonprofits in low nonprofit share areas, (2) endogeneity between nonprofit share and wage differentials or (3) lower quality workers in low nonprofit share areas. In fact, I find that nonprofit workers in low nonprofit share areas produce *higher* quality output (despite being paid less). This is consistent with the impliciations of a simple model I present in section 3.4; in particular, by maintaining lower wages nonprofits attract only workers who are "motivated" and who therefore supply higher effort than is required of them.

Thus, consistent with my theoretical predictions, I provide evidence that nonprofit wages are lowest (and nonprofit quality highest) relative to for-profits when nonprofit share (NPS) is low. Based on my model, I argue that this is driven by nonprofits' ability to hire motivated types exclusively when nonprofit share is low. Using data from the National Longitudinal Survey of Youth (NLSY), I conclude the empirical analysis by providing suggestive evidence that nonprofit workers in low nonprofit share areas are indeed a different "type" of worker. Low-NPS nonprofit workers report higher job satisfaction and are more likely to engage in prosocial activities outside of their job. The NLSY data, which includes much richer worker-level data, also allows me to rule out other possible alternative explanations for my results. For instance, I can rule out that the results are entirely explained by differences in benefits, differences in cognitive abilities of workers, and differences in firm size (which, according to existing empirical work, is also shown to impact wages.)

The remainder of the paper proceeds as follows: In section 3.2, I review in more detail existing research on nonprofit wage differentials. In sections 3.3 and 3.4, I discuss the predictions that I test empirically, first in a general way (section 3.3) and then with more precision using a simple model (section 3.4). In section 3.5, I describe the general empirical approach used throughout and apply the theoretical predictions to the environment studied. Section 3.6 reports the economy-wide analysis; I show that nonprofit share of labor indeed appears to be an important determinant of the existence of wage differences across sectors (primarily for highly educated workers.) In section 3.7, I use a richer dataset from a particular industry to rule out concerns and alternative explanations that cannot be addressed using the economy-wide data. Section 3.8 provides evidence to suggest that low-NPS nonprofits attract a different "type" of worker. Section 3.9 concludes.

3.2 RELATED LITERATURE

The main challenge in determining whether labor donation theory is a reasonable model of nonprofit labor is finding the appropriate for-profit workers to compare nonprofit wages against. That is, if what appears to be a "nonprofit" wage differential is in fact driven by differences in industry composition across the sectors, then all that exists is an industry wage differential. Two main strategies have been employed to account for the differences in industry composition across the two sectors: (1) comparing workers within a particular industry/occupation and (2) including detailed industry and occupation fixed effects in economy-wide data. However, the conclusions that result from each of these strategies have been inconsistent.

The earliest assessments of labor donation theory are typically of the first variety, in part because the data necessary to (properly) employ the second strategy did not yet exist. For instance, Weisbrod (1983) compares public interest lawyers to private lawyers and demonstrates that, controlling for observable characteristics, public interest lawyers earn significantly less. Moreover, he provides evidence from survey data that suggests that these lawyers actively select into the lower-paying field -- public-interest lawyers report being fully aware of the potential earnings they have lost and almost uniformly indicate that these losses are "worth it." Frank (1996) provides a similar result. However, using the same data as Weisbrod but with a different econometric approach, Goddeeris (1988) finds little evidence of a pay gap between public interest and private lawyers. Preston (1988), in a comparison of nonprofit and for-profit daycare center workers, finds no wage differential in non-federally regulated centers (and a *positive* nonprofit wage differential in centers that are regulated.) However, the absence of a (negative) wage differential is perhaps unsurprising in this context as more recent research demonstrates that we primarily observe differentials amongst highly educated and "white-collar" workers (Leete, 2001), a result which is found in this paper as well. Borjas et al. (1983) and Holtmann and Idson (1993) find little evidence of a pay gap between wages in nonprofit and for-profit nursing homes.

Thus, amongst the papers that attempt to determine whether the implications of labor donation theory hold in specific contexts, no conclusion has been reached. It remains unclear as to whether these divergent results stem from particular (and not yet well defined) characteristics of the industries that have been considered or if this is suggestive that, generally speaking, the implications of labor donation theory are not widely applicable. Thus, more recent research has examined economy-wide data to determine whether labor donation theory is meaningful at a more general level. An early attempt to do so indeed finds a significant nonprofit wage differential, but suffers from a lack of quality data (Preston, 1989); namely, Preston uses data from the Current Population Survey, which had not yet started asking respondents to identify whether they were nonprofit workers.

Ruhm and Borkoski (2003) also use data from the Current Population Survey to examine whether differentials exist on an economy-wide level, but are able to specifically identify nonprofit workers. They find a small negative wage differential that is not significantly different than zero. Leete (2001), using 1990 Census microdata, employs a full range of industry and occupation fixed effects. While she too fails to detect an economy-wide wage differential, she does find large and significant wage differentials in particular industries. Narcy (2011) examines wages across sectors in France and finds a significant negative wage-differential at an economy-wide level.

Although attempts to test the *implications* of the labor donation hypothesis (by searching for wage gaps) have led to inconclusive results, there is accumulating *direct* evidence of differences in workers and willingness to donate labor across sectors. In a 1977 survey, nonprofit workers were more likely to report that their work is more important to them than the money they earn (Mirvis and Hackett, 1983). Using data from the National Longitudinal Study of Youth, Benz (2005) documents higher levels of job satisfaction amongst nonprofit employees. Lanfranchi et al. (2010) find that nonprofit workers' "ideal number of hours worked" is higher than that of for-profit workers and that they are willing to receive less compensation for additional hours worked. Gregg et al. (2011) show that workers who are more willing to "donate labor" (as measured by their willingness to engage in unpaid overtime work) are indeed more likely to sort into the nonprofit sector. Serra et al. (2011) obtain survey- and experimental-based proxies of prosocial motivations and find that these measures are predictive of health professionals' selection into the nonprofit sector.

Thus, there appears to be some evidence of workers donating their labor, though the circumstances under which this leads to wage differentials is unclear. The only potential consensus to draw is that the existence and magnitude of nonprofit wage differentials depends heavily on the particular industry and/or occupation in question. This is demonstrated most clearly by Leete (2001). Yet, it remains unclear why we would observe a wage differential in some industries but not in others and also what factors are important in determining which industries are impacted.

My findings are also closely related to the more general literature on compensating differentials. Any job characteristic that is desirable to the marginal worker (such as the opportunity to support a nonprofit's mission or a low risk of being injured on the job) might be expected to generate lower wages (Rosen, 1986). However, like the literature on nonprofit wages, attempts to test the theory of compensating differentials has historically led to mixed results (Brown, 1980). One way to view this paper, then, is an attempt to locate the marginal worker (using across-locality variation in nonprofit share) and determine whether she suffers a wage penalty to compensate for the opportunity to work for her preferred job, rather than asking whether workers *on average* suffer a wage penalty. However, this paper is distinguished from much of the compensating wage differentials literature in an important sense; the workers attracted to nonprofits (despite low wages) differ not only in their willingness to accept lower wages but also – according to my argument – in their willingness to supply more effort once hired. That is, the wage gap is not merely a way to equalize utility across jobs, but also serves as a screening device to achieve the optimal match between nonprofit firms and motivated workers.

3.3 GENERAL HYPOTHESES & CONTRIBUTION TO THE LITERATURE

With the existing literature on nonprofit wage differentials in mind, I reconsider the labor donation hypothesis focusing more specifically on the circumstances under which we should expect nonprofit firms to offer lower wages. In doing so, I hope to provide a better understanding of why nonprofit wage differentials are observed in some cases but not others, particularly given recent direect evidence that nonprofit workers are willing to donate thier labor. Broadly speaking, the claims I make are the following:

1. A wage differential is observed for industries with low nonprofit share of labor relative to the for-profit sector.

2. For industries with large nonprofit share of labor, the difference between for-profit and nonprofit wages approaches zero.

The intuition behind these claims (which is described with greater precision in a simple model in the next section) begins with the assumption that there are in fact some workers who are willing to donate their labor to a nonprofit firm. (Throughout, I refer to this set of workers as "motivated.") However, the presence of motivated workers alone does not guarantee that we would observe wage differentials. Specifically, it must be in nonprofit firms' interest to maintain wages below the market wage. Of course, if workers are willing to accept lower wages, then offering the lowest wage possible minimizes costs. (Additionally, if nonprofit firms have a preference for motivated workers, then they can guarantee that only these workers will apply by setting a wage lower than the for-profit wage.)

However, this is only possible if the total demand for labor in the nonprofit sector does not exceed the number of motivated workers. Otherwise, the nonprofit firm is forced to attract standard workers; thus, the nonprofit firm must offer the for-profit wage and, because motivation is not observable, it must pay this wage to all workers (including those who are willing to work for less). As a result, whether a nonprofit wage differential is observed or not in an industry depends crucially on whether the nonprofit share of labor in that industry exceeds the proportion of motivated workers.

The above argument requires that the "motivated" worker receives *some* form of nonmonetary benefit from working at a nonprofit firm. The previous literature on labor donation theory offers a number of reasons why this might be true. Workers may receive "warm glow" or "moral satisfaction" from contributing to the production of a public good (Preston (1989), Frank (1996)). Rose-Ackerman (1996) suggests that committed workers may be easier to attract because "the lack of equity holders is a signal to employees that their selflessness is not enriching someone else." Hansmann (1980) proposes a slightly different motivation: Even if nonprofits and for-profits in an industry produce relatively similar goods, nonprofits' nondistribution constraint reduces the incentive to deviate from the promised level of quality. Thus, workers who care about the quality of service provided at their firm, may be willing to sacrifice wages to work in a nonprofit. Still other motivations are possible: Some may prefer the work environment offered by nonprofits; there is some evidence that nonprofit jobs offer greater autonomy, variety, and challenge than for-profit jobs (Mirvis & Hacket, 1983). Alternatively, individuals may be motivated to work at a nonprofit to signal to themselves or others that they are the type of person that would do so (Benabou & Tirole, 2006).

3.4 MODEL & RESULTING PREDICTIONS

In this section, I offer a simple model to add precision and more carefully consider the argument made above. The model builds upon a literature on signaling and screening of worker motivations (Delfgaauw & Dur, 2007; Heyes, 2005), which suggests that, if some workers are motivated to work in a particular job, but motivation is not observable, firms might use their offered wage as a screening device. By setting a wage lower than the reservation wage of standard workers, they are guaranteed to attract only motivated workers who receive additional nonmonetary utility from working for the firm. Here I simplify and adapt the model of Delfgaauw and Dur (2007) to understand the implications of this model in the context of nonprofit wage-setting.

3.4.1 Basic environment

Suppose there are N workers who choose between job offers at a nonprofit firm (NP) or a for-profit firm (FP). Job offers at each firm are defined solely by wage; I assume required effort is identical across the two jobs. FP serves as an outside option for workers not willing to work at NP; I therefore assume that the FP wage is the lowest wage that standard workers are willing to accept. NP chooses a wage to minimize costs.

All workers choose a job (NP or FP) and effort level to maximize $u(x)=w_j - c(x)$, where w_j is the wage in firm j, and $c(x)$ is the cost of effort. However, w_j is only received if $x \geq e$, where e is the required level of effort at either NP or FP. I assume that there are two types of workers: S

(standard) and M (motivated), where the proportion of M workers is P_M (with $0 < P_M < 1$). The utility standard workers receive is given by $u_S(x|FP) = w_{FP} - x^2$ regardless of employer. Motivated workers, to some extent, derive positive utility from offering effort to the firm. Thus, motivated workers' utility depends on the firm that they work for:

- If employed by FP: $u_M(x|FP) = w_{FP} - x^2$
- If employed by NP: $u_M(x|NP) = w_{FP} - (x - m)^2$

Motivation is captured by the incorporating m into the cost-of-effort function. Under this specification, utility is not strictly decreasing in effort and workers have some preferred level of effort that is greater than zero.

3.4.2 Worker Behavior

FP firms offer the minimum acceptable wage ($w_{FP} = e^2$), so workers never prefer to not work. Standard workers choose jobs based solely on wages. Thus, they choose FP so long as $w_{FP} > w_{NP}$ and supply minimum effort in either job.[31] Motivated workers supply minimal effort e in FP but supply effort $m > e$ in NP. Thus, they choose NP when $w_{NP} \geq w_{FP} - e^2$.

3.4.3 Nonprofit firm behavior

The nonprofit firm aims to minimize labor costs while producing the quantity, q, that is demanded by the market. That is, the firm chooses a wage to solve the cost minimization problem: $min_w[w_{NP}\, n_{NP}]$ such that $q = k\, n_{NP}\, \underline{e}(w)$ where n_{NP} is the number of workers in NP, k is the marginal product of a unit of effort, and $\underline{e}(w)$ is the average effort provided by NP workers. Average effort $\underline{e}(w)$ depends on the type of workers that sort into NP based on offered wages:

---

[31] We will assume that if wages are equal across jobs, standard workers apply for an NP job and return to FP if rejected (this is purely for convenience and does not impact the basic message of the argument).

$$\underline{e}(w) = \begin{cases} 0, & w_{NP} < w_{FP} - e^2 \\ m, & w_{NP} \in [w_{FP} - e^2, w_{FP}] \\ P_M m + (1 - P_M)e, & w_{NP} \geq w_{FP} \end{cases}$$

Thus, NP will never set $w_{NP} < w_{FP} - e^2$ or $w_{NP} > w_{FP}$. Because $e < m$, average effort is highest when $w_{FP} - e^2 \leq w_{NP} < w_{FP}$. However, the NP firm can only produce q and set a wage in this region if there are enough motivated workers to do so; that is: $n_{NP}/N \leq P_M$. If the necessary share of nonprofit labor exceeds the proportion of motivated workers – $n_{NP}/N > P_M$ – the firm is forced to offer a wage greater than or equal to that of FP to attract standard workers and produce q.

3.4.4 Resulting predictions

NP minimizes costs by choosing $w_{NP} = w_{FP} - e^2$ when $n_{NP}/N \leq P_M$ and $w_{NP} = w_{FP}$ otherwise. With $w_{NP} = w_{FP} - e^2$, only motivated types sort into the nonprofit sector. Otherwise, both types may do so. This yields two predictions:

1. Thus, a nonprofit wage differential emerges only when nonprofit labor share is relatively low and does not exceed the proportion of motivated workers in society. As nonprofit share increases, the gap between wages across the two sectors shrinks.
2. Average effort is higher in nonprofits when nonprofit share is low. As nonprofit share increases, the gap between effort across the two sectors shrinks.

For the most part, the empirical analysis will address the first prediction. However, in Section 3.7 (where I have sufficiently detailed data to assess the quality of workers' output) I provide evidence in favor of the second prediction.

3.5 EMPIRICAL APPROACH & HYPOTHESES

Given the predictions from Section 3.4, we should observe nonprofit wage differentials in industries with relatively low nonprofit labor shares (relative to for-profits in the same industry).

For an initial sense of whether this prediction is reasonable, I plot nonprofit share for each industry included in Leete's (2001) analysis against her estimated wage differentials (Figure 3.1). Leete does not discuss the relationship between nonprofit share and her estimated wage gaps, but Figure 3.1 shows that her results are consistent with the argument that negative wage gaps should appear more often when nonprofit share is low.

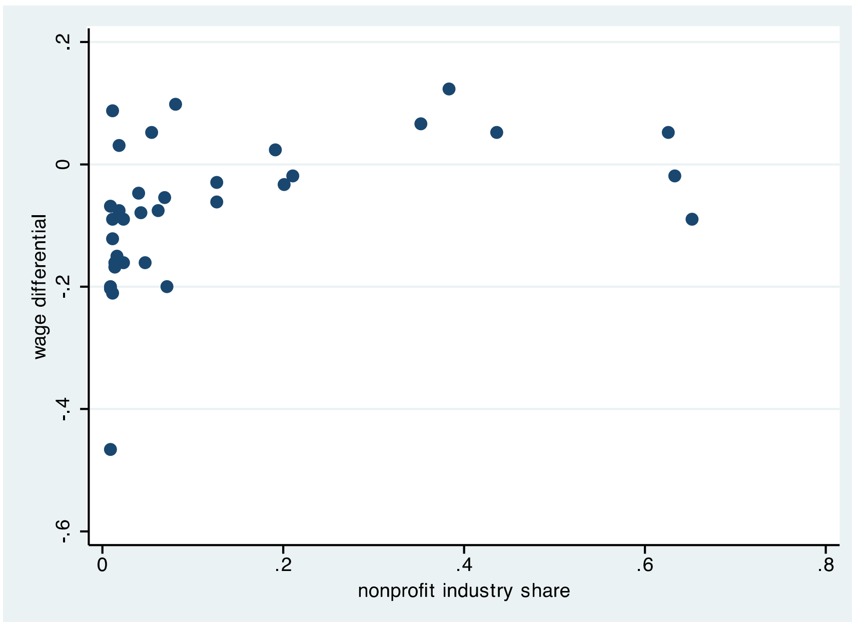

Figure 3.1. Relationship between estimated wage differential and nonprofit share from Leete (2001)

(Each point represents a particular industry)

However, because the industries being compared in Figure 3.1 are very different from one another, we should be hesitant to draw inferences from this simple relationship; the relationship could stem from other characteristics of industries that lead them to simultaneously display low nonprofit shares and larger wage differentials. This problem is of course part of the reason that there is little consensus in the existing literature that examines very specific contexts; namely, it is unclear whether to interpret the absence of a wage differential in some contexts as evidence against labor donation theory generally speaking or as simply stemming from some specific characteristics of that particular context.

To avoid this problem, I will instead compare wage differentials *within* industries but *across* localities. This will be accomplished by including a full range of industry fixed effects.

For instance, rather than comparing the wage differential amongst radio workers (an industry with low nonprofit labor share) to that of hospital workers (an industry with very high nonprofit labor share), I compare the wage differential of radio workers in a locality with *low* radio-worker nonprofit share to the differential amongst radio workers in a *different* locality with *high* nonprofit labor share. If the argument above is correct, then, for a particular industry, nonprofit wage differentials should be largest amongst localities with low nonprofit share and wages should be roughly equal in localities with high nonprofit shares.[32]

To summarize then, the broad explanation I offer for variance in differentials across industries is that, in industries with low nonprofit share, there are enough motivated workers in society (and within the particular industry in question) such that the nonprofit sector can rely only on those workers and therefore does not need to match the for-profit wage in an attempt to attract standard workers. However, for industries with large nonprofit shares (like hospitals) all of the motivated workers have already been exploited and firms must raise wages to attract other workers. Thus, wage differentials are only observed in industries with low nonprofit shares. The empirical hypotheses I will test to assess this claim are that, within industries, wage differentials will exist in localities with low industry/locality-specific nonprofit share, but as nonprofit share increases, the differential will be eliminated.

[32] Note that this argument assumes some limitations on mobility. Namely, we might think that if moving is costless, motivated workers would locate to areas where the nonprofit wage differential has been eliminated where they can obtain both higher pay and the nonmonetary benefits of working for a nonprofit. However, keep in mind that in the explanation provided above, when nonprofit share is low and a wage differential exists, motivated workers are only competing with other motivated to obtain a particular position. This is no longer true when wages are equal across sectors, in which case all workers are competing for the nonprofit job. This barrier to entry to nonprofit jobs in areas with high nonprofit shares combined with some mobility cost makes it reasonable to assume that, to some degree (and at least in the short-run), tied to their current location. Moreover, insofar as mobility does pose a threat to this empirical strategy, it would only lead to an underestimate of the true effects.

3.6 ECONOMY-WIDE ANALYSIS

3.6.1 Data and estimation approach

I first test the claim that nonprofit wage differentials depend on the nonprofit share of labor using microdata from the 5%-sample of the 2000 Census. Specifically, I construct nonprofit shares by industry at the Super-PUMA level.[33] That is, for a particular industry *i* and Super-PUMA *s*, nonprofit share (NPS) is constructed as the sum of observed nonprofit workers divided by the sum of all observed workers in the industry/locality:

$$NPS_{is} = \frac{(\text{total nonprofit employment})_{is}}{(\text{total nonprofit employment})_{is} + (\text{total for} - \text{profit employment})_{is}}$$

Because we are exclusively interested in labor markets where workers can sort into either nonprofit or for-profit jobs, I omit industry-locality groupings with nonprofit share equal to 0 or 1 – that is, labor markets totally dominated by either nonprofits or for-profits. Moreover, to avoid unreliable NPS measures I omit industry-locality groupings with less than 100 total workers, though results are similar without this restriction.

Restricting attention to full-time workers in the nonprofit and for-profit sectors, the main estimating equation is given by:

$$ln(wage) = \alpha + \beta_1 \text{nonprofit} + \beta_2(\text{nonprofit} \times \text{NPS}) + \beta_3 \text{NPS} + [\text{controls/FEs}]$$

[33] A PUMA, or Public Use Microdata Area, is a geographic grouping with a population of at least 100,000 constructed by the Census Bureau to ensure confidentiality of respondents. A Super-PUMA is a grouping of several PUMAs and has a total population of at least 400,000; Super-PUMAs are therefore relatively uniform in population, which is one reason that I use Super-PUMAs to estimate nonprofit share rather than, for instance, counties or metropolitan areas where there is great variance in population across localities. I use Super-PUMAs rather than PUMAs to increase the precision of the constructed nonprofit share measure. PUMAs are arguably better representations of "localities" but, being much smaller, counts of workers in particular industries would be a much noisier representation of the actual number of workers per industry in the area. This trade-off is not faced in the industry-specific analysis of the next section. Super-PUMAs are also more uniform in size than other geographic concepts (e.g., counties and metropolitan areas) which is important for maintaining a constant degree of precision in the constructed nonprofit share across localities.

where “nonprofit” is an indicator variable equal to 1 if the worker works for a nonprofit firm and “nonprofit X NPS” is the interaction of nonprofit employment and industry-locality specific nonprofit share.

The “wage” variable is constructed by dividing an individual’s total pre-tax wage and salary income for the preceding year, as reported in the Census microdata, by the number of hours of they worked. The income reported in the Census includes not just wages and salary, but also commissions, cash bonuses, tips, and other money income received from an employer.

In all specifications, I include industry and occupation fixed effects, as well as either state or super-PUMA fixed effects. The industry fixed effects control for any industry-specific features and therefore ensure that we are comparing wage differentials within particular industries but across localities. I also include a variety of controls such as education, potential experience (calculated as *age - years of education - 6*), gender, marital status, and race. To attempt to better capture true (but unobserved) workforce experience by interacting gender, marital status, and “potential experience” to allow for, for instance, women’s time spent out of the workforce for maternity. Throughout this section, standard errors are clustered at the Super-PUMA level.

The nonprofit wage differential, controlling for individual-level characteristics denoted *X*, is given by:

$$E[ln(wage)|\text{nonprofit} = 1, X] - E[ln(earnings)|\text{nonprofit} = 0, X] = \beta_1 + \beta_2\text{NPS}$$

If the argument from previous sections is correct, then we should observe a large negative nonprofit wage differential when NPS is low. However, this differential should become less negative as NPS increases. Thus, we would expect to find that $\beta_1<0$ and $\beta_2>0$.

It is worth noting potential empirical concerns before moving on. A common concern in assessing nonprofit wage differentials is nonrandom selection into nonprofit sector: perhaps workers in the nonprofit sector are of unobservably lower quality and would earn lower wages in either sector. This concern is minimized here in two ways: first, while this is a serious concern, this argument cannot immediately explain why nonprofit wage gaps would vary with nonprofit share (which I find that they do.) Second, in later sections I use richer data to demonstrate that (a) workers do higher quality work in low nonprofit share areas and (b) my results survive even after controlling for cognitive ability (as measured by Armed Services Vocational Aptitude

Battery scores in the National Longitudinal Survey of Youth.) Another potential concern is endogeneity of wages and nonprofit share. The Census data do not allow me to deal with this, but in Section 3.7, I use the much richer nursing home data to minimize this concern by using two instrumental variables for nonprofit share.

3.6.2 Results

3.6.2.1 Main results

Table 3.1 presents the estimation of the main specification, with and without Super-PUMA fixed effects. Negative nonprofit wage differentials exist when nonprofit share is low as can be seen from the significant and negative "nonprofit" coefficient. With nonprofit share close to zero, nonprofit workers earn 3% to 4% less than for-profit workers. This wage differential decreases as nonprofit share increases, as can be seen from the significant and *positive* "Nonprofit X NP share" coefficient.

Table 3.1. Baseline specifications

VARIABLES	(1) ln(Wage)	(2) ln(Wage)
Nonprofit	-0.0357***	-0.0273***
	(0.00471)	(0.00408)
Nonprofit X NP share	0.136***	0.126***
	(0.0110)	(0.00955)
NP share	-0.0573**	-0.0668***
	(0.0259)	(0.0197)
State FEs	X	
Super-PUMA FEs		X
Observations	899,124	899,124
R-squared	0.434	0.450

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors (clustered at super-PUMA level) in parentheses. Both specifications include additional controls as noted in the text.

However, turning to Table 3.2 which splits the sample into "high education" and "low education" groups – with "high education" defined as individuals holding a bachelor's degree or higher – we see that the predicted relationship between wage gaps and nonprofit share is much stronger for highly educated individuals. Indeed, while we observe the hypothesized signs on the coefficients in Table 3.1 (which includes high and low education groups), when we restrict our attention to highly educated individuals (columns 3 and 4 of Table 3.2) we observe both the hypothesized signs and much larger magnitudes. Highly educated workers earn between 8% and 10% less than their for-profit counterparts when nonprofit share is low; again, as nonprofit share increases this wage gap is eliminated.

Given the model and intuition behind the testable predictions, the fact that this effect is largely restricted to highly educated individuals is not surprising. The argument made in previous sections depends heavily on the idea that workers are tied to a particular industry and sort to for-profits or nonprofits *within* that industry. This assumption is more true of highly educated – and therefore more specialized – individuals. For this reason, I focus on highly educated workers for the remainder of the paper.

Table 3.2. Baseline specifications split by education group

VARIABLES	(1) Log(Wage)	(2) Log(Wage)	(3) Log(Wage)	(4) Log(Wage)
Nonprofit	-0.00303	-9.20e-05	-0.0959***	-0.0830***
	(0.00475)	(0.00449)	(0.00710)	(0.00639)
Nonprofit X NP share	0.0736***	0.0764***	0.209***	0.192***
	(0.0121)	(0.0115)	(0.0152)	(0.0136)
NP share	-0.0668***	-0.0395**	-0.0792**	-0.116***
	(0.0256)	(0.0172)	(0.0333)	(0.0316)
Educ. group	Low	Low	High	High
State FEs	X		X	
Super-PUMA FEs		X		X
Observations	564,020	564,020	335,104	335,104
R-squared	0.340	0.363	0.320	0.335

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors (clustered at super-PUMA level) in parentheses. All specifications include additional controls as noted in the text.

To provide a clearer sense of the impact of within-industry variance in nonprofit share, Table 3.3 repeats the estimation of the baseline specification (with super-PUMA FEs and high education workers only), but estimating wage gaps for one industry at a time. I focus on a set of industries which are prominent in the nonprofit sector and/or are often discussed in the previous literature on nonprofit wage differentials. In the next section, I examine a particular industry (nursing homes) in much greater detail.

We see that, with the exception of the hospital industry, within each of these industries the pattern is generally consistent with that of the main results – the nonprofit wage differential is initially negative, but shrinks as nonprofit share increases. Moreover, it is not surprising that hospitals serve as the exception here given that, unlike the other industries, nonprofit share is almost always relatively high within a locality. Thus, the only observations of hospitals in the dataset are observations wherein, according to my argument, all available motivated workers have been hired.

Table 3.3. Wage gaps within particular industries

	(1)	(2)	(3)
Industry:	Clinics	Hospitals	Research/Dev.
Nonprofit	-0.0535*	0.0272	-0.234***
	(0.0319)	(0.0262)	(0.0486)
Nonprofit X NPS	0.236	-0.00275	0.413***
	(0.207)	(0.0567)	(0.145)
Super-PUMA FEs	X	X	X
Observations	13,200	51,024	5,286
R-squared	0.454	0.286	0.367

	(4)	(5)	(6)
Industry:	Nursing homes	Media	Legal services
Nonprofit	-0.0700*	-0.227***	-0.485***
	(0.0400)	(0.0330)	(0.151)
Nonprofit X NPS	0.123	0.584*	0.457
	(0.171)	(0.328)	(2.373)
Super-PUMA FEs	X	X	X
Observations	6,562	13,141	1,454
R-squared	0.426	0.330	0.322

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors (clustered at super-PUMA level) in parentheses. All specifications include additional controls as noted in the text.

3.6.2.2 Do results vary by "collectiveness" of nonprofits?

In Section 3.3, I offered two potential explanations for labor donations from the literature. On the one hand, many nonprofits produce public or collective goods so workers may receive "warm glow" from contributing to that process. Elsewhere, as in medical-related industries, nonprofits produce essentially the same private goods or services as their for-profit counterparts, but they often produce higher quality goods (or have less incentive to deviate from the promised level of quality when quality is difficult to monitor.) To borrow Hansmann's (1980) language, a "craftsmanlike" worker – motivated by the *craft* of the job, rather than solely wages – may prefer a nonprofit because of the opportunity to participate in high-quality production.

Are the preceding results driven by one of these broad categories of nonprofits? To assess this, I split the sample using Weisbrod's index of "collectiveness" (1991). The collectiveness index is constructed as the fraction revenue within an industry that comes from public sources (private donations and government grants)[34]. Industries with high collectiveness measures receive most of their revenue from public sources and are more likely to fall into the first category discussed above: nonprofits that produce public or collective goods. Industries with lower collectiveness measures rely more on sales for revenue and as such are more similar to for-profits. Thus, low collectiveness measures are associated with the second broad category: industries where non-profits produce similar, but higher quality, goods relative to for-profits.

I split the sample into a "high" and "low" collectiveness group. An industry is included in that "high collectiveness" if its collectiveness measure is higher above the median level of collectiveness (and is included in the "low" group otherwise.) Table 3.4 reports the results of the main specification from before for each of these groups. While the magnitude of the estimated coefficients is somewhat depressed in the "low collectiveness" group, we see that the qualitative pattern observed in the baseline results holds in both cases. In the "low collectiveness" group, nonprofit workers earn 6% less than for-profit workers in low-NPS areas; in the "high collectiveness" group, nonprofit workers earn 14% less in low-NPS areas. In both cases, these wage gaps are eliminated as nonprofit share increases.

[34] I used the year 2000 "Core Files" from the National Center for Charitable Statistics to construct each industry's collectiveness measure. The Core Files collect the most essential variables from the universe of nonprofits' tax returns for a given year, which – importantly – includes "total revenue" and "revenue from public sources."

Table 3.4. Wage gaps, sample split by collectiveness

VARIABLES	(1) ***Collectiveness: Low*** ln(Wage)	(2) ***Collectiveness: High*** ln(Wage)
Nonprofit	-0.0569***	-0.138***
	(0.00930)	(0.0106)
Nonprofit X NP Share	0.174***	0.245***
	(0.0248)	(0.0192)
NP Share	-0.146***	-0.0931**
	(0.0422)	(0.0363)
Super-PUMA FEs	X	X
Observations	151,369	143,295
R-squared	0.304	0.350

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors (clustered at super-PUMA level) in parentheses. All specifications include additional controls as noted in the text.

3.6.2.3 Are wages equalized when nonprofit share is high?

Finally, while the preceding results document the predicted *directional* impact of nonprofit share on the wage gap, the theoretical predictions are somewhat more precise. When nonprofit share is low, we should see a wage gap; this has been documented. As nonprofit share increases, this gap should shrink (which is also documented) but – more specifically – nonprofit share should eventually reach a point where the wage differential is (a) constant in nonprofit share and (b) close to zero. I extend the state-fixed effects specifications and use two methods to assess these predictions: (1) a spline regression and (2) a cubic specification, allowing for a nonlinear relationship between wages gaps and nonprofit share.

In the spline regression, I allow the marginal effect of nonprofit share to vary within each quartile of nonprofit share. This provides a clearer sense of how the relationship between nonprofit share and wages vary across the domain of nonprofit share. Results are reported in Table 3.5. Again, we see a nonprofit wage gap of roughly 10% when nonprofit share is close to zero. In fact, moving into the (very small) first quartile the wage gap becomes even larger.

Consistent with the predictions, as nonprofit share increases (beyond the first quartile) we see that the marginal impact of nonprofit share decreases until it is essentially zero in the fourth quartile.

Table 3.5. Spline regression estimates of the impact of NP Share by quartile

	NPS quartile range	ln(Wage)
Nonprofit	*(0)*	-0.103**
		(0.0471)
Nonprofit X NP Share (1st quartile)	*0-0.02*	-0.702
		(2.746)
Nonprofit X NP Share (2nd quartile)	*0.02-0.07*	1.244***
		(0.378)
Nonprofit X NP Share (3rd quartile)	*0.07-0.43*	0.199***
		(0.0322)
Nonprofit X NP Share (4th quartile)	*0.43-1*	-0.0211
		(0.0395)
Observations		335,104
R-squared		0.335

*** p<0.01, ** p<0.05, * p<0.1

Robust standard errors (clustered at super-PUMA level) in parentheses. All specifications include additional controls as noted in the text.

For the cubic specification, I estimate an extension of the baseline specification:

$$
\begin{aligned}
ln(earnings) \\
= \alpha + \beta_1 \text{nonprofit} + \beta_2(\text{nonprofit} \times \text{NPS}) + \beta_3 \text{NPS} + \beta_4(\text{nonprofit} \times \text{NPS}^2) + \beta_5 \text{NPS}^2 \\
+ \beta_6(\text{nonprofit} \times \text{NPS}^3) + \beta_7 \text{NPS}^3 + [\text{controls/FEs}]
\end{aligned}
$$

The resulting wage differential is plotted in Figure 3.2, with nonprofit share along the x-axis and wage differential on the y-axis. With low nonprofit share, highly educated nonprofit workers earn roughly 12% less than their for-profit counterparts. However, this differential shrinks as nonprofit share increases until nonprofit share reaches a certain point, after which there is essentially no difference in wages across the two sectors. Taken together, the spline and cubic specifications suggest that (1) wages are equalized when nonprofit share is high and (2) beyond that point, the wage gap does not vary with nonprofit share.

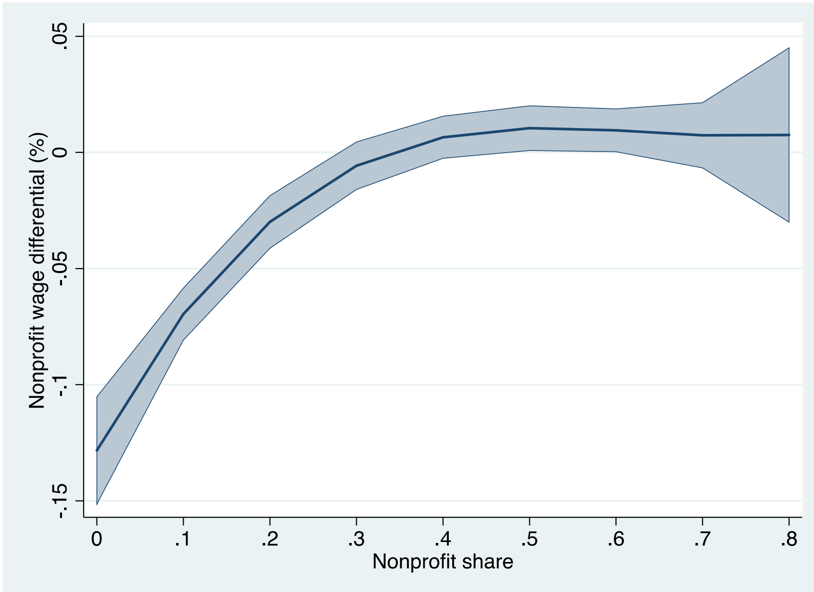

Figure 3.2. Wage differential as a function of nonprofit share

(Shaded area represents 95% confidence interval)

3.7 NURSING HOME INDUSTRY ANALYSIS

The previous section demonstrates at an economy-wide level that the existence of a nonprofit wage differential depends heavily on the nonprofit share of labor. In particular, there exist significant negative wage differentials within industries (amongst highly educated workers) in localities with relatively low nonprofit share. These differentials are diminished in localities with higher nonprofit shares. These results are in line with what we would expect to find based on the explanation put forth in the previous section: nonprofit wage differentials exist as long as there are enough "motivated" workers available to satisfy labor demand, after which firms must raise wages to draw in additional workers.

However, based on limitations of the Census data, there are several alternative explanations for the empirical results that have not been accounted for. Are nonprofit workers in low-nonprofit share areas simply lower quality, and their wages reflect this? Are the results driven by differences in the competitive environment faced by low-NPS versus high-NPS nonprofits? Alternatively, the results might be driven by differences in government attitudes towards nonprofits across localities.

To address these concerns, I turn to a similar but more detailed analysis of a particular industry – namely, the nursing home industry – with much richer data available to better account for these issues. I employ a firm-level dataset from the United States Department of Health & Human Services that contains information about every nursing home in the country that is registered with Medicare or Medicaid, thereby capturing 95% of nursing homes.[35] In the data, for each nursing home I observe: nonprofit/for-profit status, location (street address), number of residents, number of beds, total labor hours per resident, and a set of quality measures.

In addition to the richness of the data, focusing on the nursing home industry is interesting in its own right as there have been a number of papers specifically examining wage differences across nonprofit and for-profit nursing homes (e.g., Borjas et al. (1983), Holtmann and Idson (1993); Ben-ner, Paulson, and Ren (2011)). The general consensus from the literature is that, after accounting for the quality of workers, there is little difference between for-profit and nonprofit nursing home pay.[36] Thus, nursing homes provide a strong test for my central claim: although there is little evidence of a wage gap in nursing homes on average, accounting for nonprofit share may reveal an underlying wage gap that would not otherwise be detected.

3.7.1 Data

In the economy-wide analysis, a rough estimate of nonprofit share was constructed from the number of nonprofit workers per industry and locality reported in the Census data, which is why the relatively large Super-PUMA was used as the main geographic unit. Given exact numbers of labor hours used for (almost) the entire universe of nursing homes, here I can construct precise nonprofit shares within more appropriate geographic groupings. In particular, the geographic concept used in this section is the "commuting zone," as defined and constructed by Tolbert et al.

[35] The data is gathered for the department's *Nursing Home Compare* website – which is designed to help individuals find and becomes informed about nursing homes in their area. The raw data used for this website is freely available at *http://data.medicare.gov*.

[36] Ben-ner, Paulson, and Ren (2011) find little evidence of a *wage* gap, but some evidence of a *higher* level of fringe benefits in nonprofit nursing homes.

(1996). Commuting zones are clusters of counties organized around particular labor markets based on information from previous Censuses about individuals' place of residence and place of work. Commuting zones can loosely be thought of as being similar to metropolitan areas, but – unlike the metropolitan statistical areas (MSAs) – they cover the entire country. Thus, in this section I construct the nonprofit share within a particular commuting zone (CZ) as:

$$NPS_{CZ} = \frac{(sum\ of\ labor\ hours\ across\ nonprofit\ nursing\ homes\ in\ CZ)}{(sum\ of\ labor\ hours\ across\ all\ nursing\ homes\ in\ CZ)}$$

The quality ratings included in the data are constructed by the Department of Health & Human Services and are based on safety inspections and a set of ten resident health measures (e.g., "Percent of residents with pressure sores."), which therefore can be interpreted as a measure of the quality of output of employees of the nursing home.[37] These measures are reported as "star ratings," where the maximum possible number of stars is 5. The quality ratings also account for the composition of the staff (e.g., number of registered nurses relative to licensed practical nurses and nurse aides). However, while safety inspections and health measures capture the activity of the workers, the staffing measures only capture *who* is hired and not the effort they provide. Thus, I modify the reported overall measure slightly to strip away the variation that stems from staffing measures. (I do, however, control for the staff composition in the empirical analysis that follows.) In particular, the overall measure accounts for staffing in the following manner: "Add one star to […] if staffing rating is four or five stars and greater than the health inspection rating; subtract one star if staffing is one star." I simply reverse this process, substracting a star if staffing is four or five and adding a star if the staffing measure is one star.

The nursing home data *does not* include any information about wages. I pair the nursing home data, which I use to construct locality-specific information such as nonprofit share, with microdata from the 2010 American Community Survey (ACS). The ACS asks many of the same questions as the Census data used in the economy-wide analysis, but only provides a 1%-sample of the United States population. The size of the sample is less critical here though, as I am not

[37] Detailed descriptions of the construction of these quality ratings can be found at http://www.cms.gov/CertificationandComplianc/13_FSQRS.asp

relying on this data to construct estimates of nonprofit shares (as was the case in the previous section.)

3.7.2 Empirical approach & baseline results

The general empirical strategy is the same here as in the previous section. I restrict my sample, drawn from the ACS data, to individuals who work in the nursing home industry and, as a baseline specification, estimate the same empirical specification as in preceding sections.

I include the same individual-level controls as before (race, education, the interactions of expereince, gender, and marital status), commuting-zone fixed effects (which capture the main effect of NPS), and occupation fixed effects. Standard errors are clustered at the commuting-zone level.

The results of this baseline specification – essentially replicating the approach from the previous section but with a single industry and a better NPS measure – are reported in Column 1 of Table 3.6. Again, as before, we find a large and significant negative wage differential when nonprofit share is low (as indicated by the "nonprofit" coefficent) that disappears as nonprofit share increases (as indicated by the postive & significant "nonprofit X NPS" coefficient.)

3.7.3 Controlling for market conditions

Aside from eliminating noise from the NPS measure, the estimation presented in column 1 is subject to the same concerns as in the economy-wide data. Thus, in the remaining columns of Table 3.6, I attempt to minimize such concerns. In particular, it is possible that the economy-wide results are driven by differences in the market conditions that low- and high-NPS nonprofits face. However, because these preceding results include locality fixed effects (Super-PUMA fixed effects in the previous section and CZ fixed effects here), "market conditions" can only impact the results insofar as they *differentially* impact the way that nonprofits and for-profits set wages.

In columns 2 and 3 of Table 3.6, I add controls for local market characteristics that were not possible to account for in the economy-wide data and – because it is only the *differential* impact of market conditions on nonprofits' wage setting that we are concerned about – these measures are interacted with whether or not the worker is employed by a nonprofit. All market-level characteristics are re-centered around their means so that we can still interpret the "Nonprofit" coefficient as the wage gap when non-profit share is zero and all other locality characteristics are at their mean.

Table 3.6. Nonprofit wage gaps in nursing homes with market-level controls

VARIABLES	(1) ln(Wage)	(2) ln(Wage)	(3) ln(Wage)
Nonprofit	-0.173***	-0.186***	-0.175***
	(0.0533)	(0.0635)	(0.0543)
Nonprofit X NPS	0.386***	0.418**	0.389***
	(0.141)	(0.180)	(0.144)
Nonprofit X (# of residents)		2.77e-07	
		(2.58e-06)	
Nonprofit X (# of for-prof. firms)		-0.000194	
		(0.00128)	
Nonprofit X (# of non-prof. firms)		0.000139	
		(0.000296)	
Nonprofit X HHI			-0.135
			(1.131)
Constant	4.222***	4.219***	4.222***
	(0.280)	(0.282)	(0.280)
Observations	2,082	2,082	2,082
R-squared	0.506	0.506	0.506

Robust standard errors (clustered at commuting zone level) in parentheses. All specifications include additional controls as noted in the text.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

In column 2 I control for directly observable measures: number of nursing-home residents in the CZ, number of for-profit nursing homes in the CZ, and number of nonprofit nursing homes in the CZ. These measures control for the general size of the market but also loosely account for the relative market concentrations of nonprofits and for-profits. For instance, it is possible for nonprofit share to be high either because there is one large nonprofit nursing home or many small nonprofit nursing homes. This difference of course would not be picked up by NPS alone, but could impact wage-setting. Similarly, in column 3, these measures are summarized into a single constructed variable: the Herfindahl-Hischmann Index for nursing homes within the CZ.

In both columns 2 and 3, we see that these market characteristics do not significantly *differentially* impact the wages of nonprofit workers. (They presumably impact wages *generally* but, again, any general effect is captured by the CZ fixed effects.) More importantly, we see that, even accounting for these characteristics, the general result from the baseline model holds: a significant and negative "nonprofit" coefficient paired with a significant and positive "nonprofit X NPS" coefficient.

3.7.4 Endogeneity of nonprofit share

There is a concern that nonprofit share and the wage differential are endogeneous. In particular, if a state or local government provides conditions which reduce costs for nonprofits to a greater degree than elsewhere, then nonprofits may have more funds available to pay workers. At the same time, a generally more nonprofit-friendly environment is also likely to increase nonprofit share. To attempt to address this, I use a two-stage least squares instrumental variables approach, taking an instrument for nonprofit share that has been used in previous literature (Grabowski & Hirth, 2003; Sloan et al., 2001). In particular, I instrument for NPS using the *growth in the elderly population* (ages 65 and above) in the first half of the decade, from 2001-2006. As Grabowski and Hirth (2003) discuss, for-profit nursing homes face lower start-up costs and can more quickly react to demand. Thus, a locality with higher recent growth in elderly population is likely to have a *lower* nonprofit share for reasons unrelated to local government's preferences towards nonprofits.

The results of this estimation are displayed in columns 1 and 2 of Table 3.7. The empirical specification includes all of the controls employed in the baseline results reported in Table 3.6. Column 1 reports the first stage result: as expected, a growth in the elderly population decreases nonprofit share. Column 2 reports the second stage. The result from the preceding section survives (and, in fact, is stronger.)

The growth in elderly population represents an increase in demand for nursing homes generally, rather than a differential increase in demand for nonprofits. Because we are interested in the *differential* impact that nonprofit share has on nonprofit wages, the elderly population instrument plausibly satisfies the exclusion restriction.

However, *differential* changes in demand for nonprofit nursing homes in particular present an additional potential source of endogeneity. If demand for nonprofit nursing home services increases in a particular locality (perhaps because nonprofit nursing homes are expected to provide higher quality), demand for labor in nonprofit nursing homes would increase as well. Thus, we might expect an increase in both nonprofit share and wages. I use a second instrumental variable to address this concern: ten-year lagged nonprofit share in *unrelated* industries within the same locality. The set of "unrelated industries" excludes any medical-related industry. Lagged unrelated-NPS caputres the fact that a locality may have had nonprofit-friendly policies in the past which attracted nonprofits, thus setting in place a nonprofit share that is determined independent of demand for nonprofit nursing homes. Results are reported in columns 3 and 4. Column 3 reports the first stage: higher lagged, unrelated NPS is associated with higher nursing home NPS. Column 4 reports the second stage estimates. Again, we observe wage gaps in low nonprofit share areas that are eliminated as nonprofit share increases.

Table 3.7. Two-stage least squares estimates to account for endogeneity of nonprofit share

	(1)	(2)	(3)	(4)
	IV: *Growth in elderly population*		**IV**: *Lagged unrelated-industry NPS*	
VARIABLES	**First stage**	**Second stage**	**First stage**	**Second stage**
Nonprofit		-0.352**		-0.298***
		(0.163)		(0.101)
Nonprofit X NPS		0.996*		0.812**
		(0.558)		(0.339)
Instrument	-0.708***		3.081***	
	(0.247)		(0.687)	
Observations	2,084	2,084	2,084	2,084
R-squared	0.890	0.503	0.903	0.505

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors (clustered at commuting zone level) in parentheses. All specifications include additional controls as noted in the text.

3.7.5 Quality of work in nonprofits

Finally, an important concern in the economy-wide anaylsis is the possibility that, in localities with low nonprofit share, nonprofit workers are less productive and the negative wage differential simply reflected this quality differential. The simple model in Section 3.4 predicts just the opposite. In this section, I take advantage of the nursing home five-star quality measures constructed by the Department of Health & Human Services and assess the quality of a nursing home's output as a function of their nonprofit status and the nonprofit share in their locality. That is, taking a nursing home as the unit of observation, I estimate the same specification as before but I take the quality measure as the dependent variable. As in previous regressions, I include commuting zone fixed effects and cluster standard errors at the commuting zone level. In one specification, I include firm-level controls: the nursing home's share of their market, registered nurse (RN) labor hours per resident, licensed practical nurse (LPN) labor hours per resident, other staff hours per resident, total residents, and ratio of total residents to total beds.

Table 3.8. Quality differentials as a function of nonprofit share

	(1)	(2)
Nonprofit	0.712***	0.386***
	(0.0820)	(0.0819)
Nonprofit X NPS	-0.797***	-0.398*
	(0.236)	(0.221)
Firm market share		-1.413
		(0.915)
RN Hours/resident		0.231***
		(0.0289)
LPN Hours/resident		-0.116***
		(0.0414)
CNA Hours/resident		0.162***
		(0.0315)
Residents		-0.004***
		(0.001)
Fraction of beds occupied		1.916***
		(0.114)
Constant	2.664***	1.147***
	(0.0120)	(0.138)
Observations	13,393	13,089
R-squared	0.117	0.182

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors (clustered at commuting zone level) in parentheses.

If the previous results were in fact driven by less productive workers in low-NPS areas, we would expect to see a similar pattern amongst these coefficients as we have observed in previous results: the "nonprofit" coefficient would be significantly negative while "nonprofit X

NPS" would be significantly positive. However, as can be seen in Table 3.8, we observe just the opposite; quality at nonprofits is significantly *higher* quality when nonprofit share is low and *decreases* amongst nonprofits in higher nonprofit share areas. That is, it is precisely the areas where nonprofit nursing home pay is *lowest* that nonprofit quality is *highest*, and as nonprofit share increases (and nonprofit pay increases) the quality of output at nonprofits moves closer to that of for-profits. This is consistent with a model – such as the motivation-screening model presented in Section 3.4 – wherein lower wages in nonprofits draw in workers who are willing to put forth greater effort than is required of them.

3.8 DO NONPROFITS IN LOW NONPROFIT SHARE AREAS ATTRACT A DIFFERENT "TYPE" OF WORKER?

Two main results emerge from the preceding sections of the paper:

1. At the labor market level, wage gaps are largest in low nonprofit share areas.
2. At the firm level, nonprofit nursing home output is highest in low nonprofit share areas.

These results are *consistent* with the suggestion that a wage gap serves as a motivation-screening device and, as a result, a *specific type* of worker is attracted to nonprofits in low nonprofit share areas. However, the preceding results do not provide direct evidence that motivation-based sorting is the mechanism driving the results. That is, due to data limitations, I provide no evidence that low-NPS nonprofit workers are observationally different than high-NPS nonprofit workers in ways that might suggest that they are more likely to be "motivated types." Using much richer worker-level data from the National Longitudinal Survey of Youth (NLSY), I provide suggestive evidence to fill this gap. I also take advantage of the NLSY to address alternative explanations remaining from previous sections.

As proxies for "motivation," I focus on two measures available in the NLSY: job satisfaction and prosocial activities (volunteering and donating). The interest in these two measures is generated by existing findings. Benz (2005) documents higher average job satisfaction amongst nonprofit employees. This could be due to characteristics of the work environment, in which case all workers benefit, or it could be an indication of higher utility "donating labor," in which case only "motivated" workers benefit. Thus, if the preceding results are indeed driven by more "standard" workers sorting into nonprofits as nonprofit share increases, then job satisfaction should be particularly high where nonprofit share is low (and only motivated workers are present.) Satisfaction should diminish in higher nonprofit share areas.

Similarly, previous researchers have found that nonprofit workers are in general more likely to donate or volunteer (Houston, 2006; Serra et al., 2011), which some take as an indicator of "public service motivation." As with job satisfaction, there are multiple possible explanations for a higher propensity to engage in prosocial activities. For instance, it is possible that nonprofit workers simply have more access to or information about donation/volunteering opportunities. If this were the only explanation, then we would expect no relationship between likelihood of engaging in prosocial activities and nonprofit share. However, if this increased likelihood were indeed an indication that some nonprofit workers are "motivated types," then here we would expect this result to be particularly strong amongst low nonprofit share workers.

3.8.1 Data & empirical approach

The data used in this section is drawn from the 2000 to 2010 waves of the 1997 National Longitudinal Survey of Youth (NLSY). The NLSY surveys a panel of individuals yearly. By 2000, respondents were between the ages of 15 and 19. By 2010, respondents were between the ages of 25 and 29. (Because I restrict the sample to workers with a bachelor's degree or higher, as in the rest of the paper, in reality the first observations are in 2002.) Thus, the sample consists of workers just sorting into nonprofit or for-profit jobs for the first time. This is ideal as the theoretical predictions speak to the impact of this job-choice decision (as opposed to, for instance, evolution of wages once employed.)

Relative to data sources used elsewhere in this paper, the NLSY includes relatively few observations and therefore very few observations of nonprofit workers. Pooling data from all of the waves between 2000 and 2010, there are 3,608 observations of employed, "highly educated" workers. Of these, 473 report working in a nonprofit. As a result, although the NLSY is a panel, I analyze the data as pooled cross-sections. In all specifications, I cluster standard errors at the individual level to account for multiple observations of each respondent.

I use the confidential geocoded version of the NLSY97 to identify respondents' county of residence, which I then map into Super-PUMAs to assign each worker a nonprofit share based on his or her industry and Super-PUMA. The construction of nonprofit share is identical to the procedure used in Section 3.6, with the exception that nonprofit share is required for every year between 2000 and 2010. As before, I use the 2000 5%-sample of Census microdata for 2000 nonprofit shares. I also construct nonprofit shares for 2006 and 2009 using pooled 3%-samples of American Community Survey data.[38] For all the remaining years of the decade, I linearly interpolate between 2000, 2006, and 2009 to construct industry/locality-specific nonprofit shares for each year.

Despite the small number of I observations, I turn to the NLSY because it includes a much richer set of worker-level variables than the data sources used in preceding sections. As proxies for "motivation," I focus on two measures available in the NLSY: job satisfaction and prosocial activities. I collapse the five-point scale job satisfaction responses into a binary variable. My constructed variable, *Satisfied*, is equal to one if respondents indicate that they like their jobs "very much" or "fairly well." *Satisfied* is equal to zero if respondents indicate that they neither like nor dislike their job or express negative feelings. As a measure of prosocial activities, I use responses from the 2005 and 2007 waves. In these two waves, respondents were asked whether they engaged in any volunteering or made any donations in the preceding year. I construct a binary variable, *Prosocial*, equal to one if respondents answer yes to either of these questions.[39]

I attempt to match the basic empirical specifications used in previous sections. I regress the outcome variable of interest (log of wage, satisfaction, or prosocial activities) on a dummy for nonprofit, nonprofit share (NPS), and the interaction of nonprofit and NPS. I also include controls that match those used in previous sections: gender, race, age, marital status, and highest degree received. Unlike in the previous data, I can control for *actual* experience in the labor market as the NLSY reports the total number of weeks an individual has worked since the age of

---

[38] These data were obtained from ipums.org. 2006 is the first year that a pooled 3% sample is available. The "pooled samples" pool three years worth of 1% samples. The 2006 sample, for instance, pools the 2005, 2006, and 2007 1% ACS samples.

[39] Respondents who report engaging in volunteer work are also asked whether the volunteer work was truly voluntary and not, for instance, court-ordered. I only count an individual as having volunteered if they engaged in voluntary volunteering.

20 and tenure with their current employer. Because of the limited number of observations, there are far fewer individuals in any given industry, occupation, or location. I therefore use broader fixed effects for each of these than in previous sections. For location, I employ state fixed effects. For industry and occupation, I collapse the three-digit codes into the broader categories listed by IPUMS.[40] However, results are generally consistent when using more detailed industry, occupation, and location fixed effects.

3.8.2 Results

I begin by demonstrating that the main wage gap result from the preceding section holds in this sample and is robust to additional controls available in the NLSY. Table 3.9 reports results, taking (log of) wage as the dependent variable. Model 1 is the baseline specification, employing controls noted above. As in Sections 6 and 7, nonprofit workers in low nonprofit share areas earn significantly less than comparable for-profit workers, but this wage gap is eliminated as nonprofit share increases. With the richness of the NLSY data, we can extend this baseline specification to rule out additional alternative explanations. Model 2 adds controls for respondents' ASVAB scores, which is commonly used in the literature as a measure of cognitive ability. The results are essentially unchanged.

Model 3 adds to the baseline specification a set of dummy variables indicating which benefits the respondent's employer provides (e.g., medical insurance, flexible work schedule, etc.). Model 4 adds to the baseline specification a control for the number of other people that work for the same employer as the respondent. Existing empirical work finds that workers in large firms earn higher wages (e.g., Troske (1999)). If market-level nonprofit share is correlated with size of nonprofit firms, then this could present an alternative explanation for preceding results.[41] Responses to questions about benefits and "number of employees" are not available for all respondents (explaining the changing number of observations), so the sample in Models 3 and

---

[40] For instance, all mining-related industries (coal mining, metal mining, etc.) are collapsed into one broader "mining" category. See: http://usa.ipums.org/usa-action/variables/IND1990#codes_tab and http://usa.ipums.org/usa-action/variables/OCC1990#codes_tab for industry groupings.

[41] I can also address this concern by adding a control for the market-level average size of nonprofit and for-profit nursing homes in Section 7. I do not report the results, but adding such a control does not impact the results.

4 slightly differ from the sample in Models 1 and 2. Nonetheless, the main result holds after adding these controls; the results do not appear to be explained by differences in benefits or the large-firm wage premium.

Table 3.9. Estimating wage gaps in the NLSY97

VARIABLES	(1) Baseline	(2) Control for ability	(3) Control for benefits	(4) Control for number of employees
Nonprofit	-0.419***	-0.412***	-0.379**	-0.255*
	(0.135)	(0.136)	(0.153)	(0.147)
Nonprofit X NPS	0.861***	0.837***	0.526	0.408
	(0.300)	(0.301)	(0.326)	(0.324)
NPS	-0.504**	-0.504**	-0.174	-0.0781
	(0.202)	(0.203)	(0.238)	(0.240)
Observations	865	865	621	538
R-squared	0.332	0.335	0.476	0.465

*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors (clustered at individual-level) in parentheses.
All specifications include controls as noted in text.

Having established that the main result holds amongst NLSY97 respondents, we now turn to differences in job satisfaction and prosocial activities. Results, taking these two measures as dependent variables, are reported in Table 3.10. Low-NPS nonprofit workers are significantly more likely to report that they enjoy their job (Column 1) and are also more likely to report engaging in prosocial activities (Column 2). These results are diminished as nonprofit share increases.

Table 3.10. Differences in proxies for motivation as a function of nonprofit share

VARIABLES	(1) Satisfied	(2) Prosocial
Nonprofit	0.225**	0.575***
	(0.0895)	(0.213)
Nonprofit X NPS	-0.251	-0.497
	(0.217)	(0.497)
NPS	0.0124	0.681
	(0.201)	(0.446)
Constant	0.479	1.317
	(0.524)	(1.148)
Observations	715	274
R-squared	0.223	0.459

*** p<0.01, ** p<0.05, * p<0.1

Robust standard errors (clustered at individual-level) in parentheses. All specifications include controls as noted in text.

3.9 CONCLUSION

Existing research attempts to determine whether the nonprofit wage gap is driven by some workers' willingness to accept lower wages (the "labor donation hypothesis") or by the composition of jobs in the nonprofit sector. Most researchers either look within a particular industry where we can be confident that jobs are relatively similar or employ detailed controls for industry, occupation, etc. This has led to inconsistent results: clear wage gaps exist in some industries but not others. The existence of a wage gap is taken as evidence of the labor donation hypothesis; the absence of a wage gap is often taken as evidence *against* labor donations.

I contribute to this literature by asking: assuming some workers *are* willing to "donate their labor," when should we *expect* to see a wage gap? I build on a point made by Preston in 1989; even when some workers are willing to donate their labor, in order for a wage differential to exist, it must be in firms' interests to *maintain* low wages. Once the number of workers demanded by nonprofits exceeds the number of "motivated" workers, firms must raise their wages. This, then yields the prediction that the existence of a nonprofit wage differential within a particular industry depends on how much labor nonprofits demand relative to for-profits, with wage differentials existing only in industries with relatively low nonprofit shares of labor.

I provide empirical evidence consistent with this suggestion. To avoid making comparisons across industries, I examine nonprofit wage differentials as a function of nonprofit share *within industries* but *across localities*, first through a detailed set of industry fixed-effects in an economy-wide analysis and then by examining the nursing home industry in more detail. Throughout, it is indeed the case that wage differentials exist when nonprofit share is low (and firms are able to draw in motivated workers) that then disappear as nonprofit share increase (with firms required to draw in standard workers.) Using a detailed firm-level dataset that encompasses 95% of nursing homes within the United States, I provide evidence that suggests that these results are *not* driven by differences in competitive environments or endogeneity of nonprofit share and nonprofit wages. I also show that in the nursing home industry, the quality of nonprofit work is *highest* when nonprofit share is low and decreases as nonprofit share increases. This result is striking as it demonstrates that the nonprofit workers being paid the least (relative to their for-profit counterparts) are also the workers who are producing the highest quality work. Moreover, using NLSY data, I show that nonprofit workers facing the largest wage gaps are also most likely to report enjoying their job.

Thus, empirically, there is clear evidence that the existence of nonprofit wage differentials depends on nonprofit share. This systematic relationship helps organize some of the divergent results in the existing literature on nonprofit wages. It had been established, for instance, that nonprofit workers in legal services earn less than for-profit workers, but workers in the hospital industry are not penalized for choosing a non-profit job. Within the framework of this paper, this is unsurprising: nonprofit share in legal services is very low, while nonprofit share in hospitals is very high.

Beyond this empirical regularity, the relationship wage gap and nonprofit share is consistent with the argument that when nonprofit labor demand is relatively low, nonprofits can maintain low wages and rely on “motivated” workers. As nonprofits’ labor demand grows, they are forced to attract standard workers as well and must raise their wages to attract them. The observed pattern of wage gaps as a function of nonprofit share provides indirect evidence of this motivation-sorting and screening mechanism. The results I present on firm-level quality, worker-level job satisfaction, and worker-level prosocial activities provide more direct evidence that a different “type” of worker is indeed sorting into nonprofits when wage gaps are large.

This paper therefore provides empirical support for recent theoretical work on sorting and screening of "motivated workers" in the labor market (Besley & Ghatak, 2005; Heyes, 2005; Delfgaauw & Dur, 2007; Ghatak & Mueller, 2011), which applies to not just the nonprofit sector, but also teaching, nursing, or any other job where workers might claim that they are "not in it for the money." I show that the sorting of motivated workers into jobs that they "care about" can have important and economically significant implications for labor market outcomes. My results suggest that raising wages may not always be a clear solution to attracting more productive workers (an idea suggested in recent theoretical work (Delfgaauw & Dur, 2010)), as I find that quality of output is highest in nonprofit nursing homes precisely when nonprofit wages are lowest relative to for-profits. This echoes recent empirical evidence that higher pay for politicians – another occupation where "mission-oriented motivation" may be important – reduces politician quality (Fisman et al., 2012). Perhaps more importantly, I build on the theoretical literature on motivated labor by demonstrating that – in practice – the impact of motivation in the labor market depends heavily on both the supply and demand sides of the market.

4.0 WALLFLOWERS: EXPERIMENTAL EVIDENCE OF REPUTATION AVOIDANCE (WITH SERA LINARDI)

4.1 INTRODUCTION

How is prosocial behavior impacted by visibility? Given the frequency with which charitable contributions are rewarded with public recognition – be it through a listing of donors in a monthly newsletter or the naming of a building in a donor's honor – fundraisers seem to believe that visibility has a positive effect. The majority of current research seems to support the hypothesis that visibility should increase giving due to a desire to acquire a positive reputation[42] -- that is, *appearing* altruistic to others (Benabou & Tirole, 2006). For instance, Andreoni and Petrie (2004) and Rege and Telle (2004) find that visibility increases voluntary contributions to a public good.[43] However, there is evidence to the contrary: Noussair and Tucker (2007) find that visibility decreases giving in a repeated public goods game. Dufwenberg and Muren (2006) find that visibility decreases giving in a dictator game. Shi (2011) finds that visibility decreases participants' willingness to donate blood. Others have found that the success of visibility depends heavily on the context or nature of the decisions being made (Alpizar et al., 2008; Gachter & Fehr, 1999; Noussair & Tucker, 2007; Soetevent, 2005).[44] If the desire to appear

[42] Throughout the paper, "reputation" refers to intrinsic reputation concerns as opposed to strategic reputation concerns. For agents in our paper (as in Benabou & Tirole, 2006), others' perception of them directly enters their utility function and therefore impacts behavior in a one-shot game. This is distinct from strategic reputation concerns that arise in repeated games. It is worth noting that a number of papers have focused on sustaining cooperation in repeated games through strategic reputation concerns and "indirect reciprocity" (e.g., Milinski & Wedekind, 2000; Seinen & Schram, 2001; Wedekind & Braithwaite, 2002; Bolton et al., 2005).

[43] See Bekkers and Wiepking (2010) for a review of research from psychology and economics.

[44] Gachter and Fehr (1999) find that contributions to a public good increase with visibility when contributors meet and interact beforehand, but not otherwise. Soetevent (2005) in a field experiment manipulates the visibility of church donations and finds a positive impact for external causes but no impact of internal fundraising. In

altruistic were a general and accurate description of individuals' preferences, we would expect to find that visibility increases prosocial behavior more consistently.

In this paper, we propose and provide evidence of a different type of response to visibility: reputation avoidance. In particular, we hypothesize that some individuals are "wallflower" types, who are averse to standing out either in a negative way *or* positive way. Why would people be averse to positive reputation? Benabou and Tirole discuss the possibility that incentives for prosocial behavior might not work if people are suspected of being motivated by incentives.[45] There is a wealth of evidence showing that *monetary* incentives reduce prosocial behavior when visible; mixed evidence on the impact of visibility may suggest that this is true for reputation incentives as well. More generally, there is evidence that people are sometimes punished for being "too good". In the context of laboratory public goods games, individuals are willing to punish contributions that are too low *or* too high (Herrmann, Thoni, & Gachter, 2008).[46] There is also theoretical precedence for the idea that individuals react to scrutiny by avoiding extreme actions: in Bernheim's model of conformity (1994) "extremists are esteemed less than centrists." Similarly, our model of wallflower preferences predicts that individuals move towards an action that signals they are average in their level of altruism. This forms a close relationship between the impact of visibility and individuals' beliefs about what others will do or have done. If a donor believes (or observes) that the "average" individual gives a high (low) amount, then by moving towards that average type visibility may have a positive (negative) impact.

Some existing work is suggestive of wallflower behavior. Linardi and McConnell (2011) find that subjects are more likely to quit volunteering for a nonprofit when they recently observed another subject leaving when highly visible. Peacey and Sanders (2012) study the decision to make a donation anonymously and find that individuals who choose extreme low *or* high donations are more likely to conceal their identity.

---

an experiment at a national park, Alpizar et al. (2008) finds that allowing solicitors to directly view visitors' donation only increases giving when donors received a small gift and not otherwise. Noussair and Tucker (2007) find that visibility decreases contributions in a repeated public goods game.

[45] In BT model where honor is positive, reputation incentives would never backfire (reduce giving). It will just cease to become effective. In our model of wallflower preferences where honor is negative, reputation incentives can backfire.

[46] Outside of prosocial behavior, trying too hard or performing well in school is sometimes greeted by disapproval by peers as is the case, for instance, in the much-discussed "acting white" phenomenon (Austen-Smith & Fryer, 2005).

To more directly identify wallflower behavior in our experiment, we exogenously manipulate both visibility and social information to study how individuals react to information high *and* low levels of giving. Participants are assigned to anonymous groups of three. In the first stage, participants decide how much to contribute to a charitable cause. Participants in the *Baseline* treatment are informed that their donations will be submitted in a sealed envelope at the end of the experiment. Participants in the *Visibility* treatment are told that their group will submit their donations in front of one another. In the second stage, participants are given an opportunity to change their donation by conditioning their contribution on every possible combination of the other two group members' donations. We find support for our predictions: participants in the Visibility treatment are more likely to condition their donations to fall *within* the range of their group members' contribution, thus avoiding standing out in a negative or positive way. As a result, visibility only has a positive impact when group members' contributions are high.

We also test for heterogeneous responses to visibility by examining differences in behavior according to familiarity with the cause and gender. Existing work suggests that (1) social information primarily impacts the donations of individuals who do not have prior experience with the organization (Shang & Croson, 2009) and (2) that "the social preferences of women are more situationally specific than those of men" (Croson and Gneezy, 2009).[47] Our main result – that visibility increases the frequency of contributions that fall between the range of others' donations– is indeed driven by women and individuals unfamiliar with the charitable cause. Visibility has an additional effect of compressing females' unconditional giving around their beliefs about average giving, suggesting that "wallflower" reputation preferences may be particularly strong among females.

We test the generalizability of our result in a small-scale field experiment; we find additional evidence of wallflower behavior, particularly amongst women. In the field experiment, participants perform real effort tasks to raise funds. Here, the modal giving is low. We find that women, *when visible*, match what they observe others give and as a result give less in this context.

[47] See Zetland and Della Giusta (2011), Mellstrom and Johannesson (2008), Lacetera and Macis (2010).

4.2 MODEL

In this section, we contrast the predictions of Benabou & Tirole's model of prosocial behavior -- which assumes that individuals seek positive reputation -- with our modification thereof, wherein individuals avoid any form of reputation. We will refer to these reputation-avoiding types as "wallflowers." Here, we merely discuss the assumptions and predictions of the model.

In Benabou & Tirole's (2006) (hereafter BT) model of prosocial behavior, agents are intrinsically motivated to be prosocial but are also concerned about what their actions imply about their level of altruism. They experience positive utility when their actions signal that their type is above average (honor) and negative utility when their type is below average (stigma). We extend this widely used model of reputation concerns by 1) linking it explicitly to social information, and 2) exploring the consequence of allowing individuals to desire neither honor nor stigma.

4.2.1 Basic model

Let an individual's intrinsic altruism be represented $v_i \sim u[0,A]$. Assume that individuals choose between N+1 discrete contribution levels $a_i \in\{0,1,2,\ldots,N\}$ with cost function $C(a_i)$ where $C'(a_i)>0$ and $C''(a_i)>0$.[48] Following BT, let $v_i a_i$ represents the intrinsic benefit of prosocial behavior for a person with altruism v_i who contributed a_i. Let $x\in\{0,1\}$ denote whether or not the individual's decision (and identity) is publicly observable.

Following BT, the basic components of utility considered in prosocial decisions are intrinsic benefit, cost of action, and reputational benefits:

$$u(a|x,v) = v_i a_i - C(a_i) + xR(a_i)$$

[48] Because any portion of the endowment that is not contributed is consumed, this assumption can also be viewed as decreasing marginal benefit of consumption.

where $R(a_i)$ represents the reputational benefits associated with choosing a_i when x>0. For individuals with standard BT preferences (who we refer to as reputation-seeking types from here on):

$$R(a_i) = E(v|a_i) - E(v)$$

That is, the reputation a reputation-seeking type obtains from choosing an action is dictated (a) by the altruism type v that can be inferred from the action and (b) how distant that type is from the average altruism type (E(v)). Note that, in either case, $R(a_i)$ is positive when $E(v|a_i) > E(v)$ and negative otherwise.

For individuals with wallflower preferences:

$$R(a_i) = -|E(v|a_i) - E(v)|$$

That is, any action that leads others to assume that the agent's type $E(v|a_i)$ is different than the average type $E(v)$ reduces the agent's utility.

Define the marginal benefit of reputation in moving from *a-1* to a as $r(a) = R(a) - R(a-1)$. For reputation-seeking types, this is equal to $r(a) = E(v|a) - E(v) - E(v|a-1) + E(v) = E(v|a) - E(v|a-1)$. Because $E(v|a) - E(v|a-1) > 0$ (higher altruism types are expected to choose higher actions), the marginal reputation of a higher contribution is always positive. This is not true for wallflowers. For wallflowers, marginal reputation is negative if $E(v|a_i) > E(v)$ and positive if $E(v|a_i) < E(v)$. That is, moving closer to the average type – from above or below – increases utility.

In summary, these preferences imply:

1. Wallflowers and reputation-seeking types behave the same in the absence of visibility.
2. Reputation-seeking types: Visibility increase prosocial behavior. This is because the marginal benefit of reputation is always positive.
3. Wallflower types: Visibility increases a wallflower's likelihood of choosing the action taken by the average altruism type and decreases the likelihood of choosing an extremely high or low action.

4.2.2 Incorporating social information

In some situations (as in the second phase of our laboratory experiment), an individual receives information about the actions of others who faced the same choice before making her own decision. We will refer to this as social information and denote it with s. If this social information is publicly available when the individual's decision is revealed, it is likely that s will affect whether the individual's contribution is perceived as "high" or "low" by others. The salient "average type" is no longer the average altruist in the population $E(v)$ but $E(v|s)$.

First note that if no social information is received, an individual's expectation of the average altruism type is simply the population average: E(v)=A/2. Now suppose that (as in our laboratory experiment) an individual observes that actions of two other individuals j and k, in other words, s=(a_j, a_k). Given this social information, $E\left(v\middle|a_j, a_k\right) = 0.5(a_j + a_k)$. Reputation concerns are $R(a_i) = E(v|a_i) - E\left(v\middle|a_j, a_k\right)$ or $R(a_i) = -\left|E(v|a_i) - E\left(v\middle|a_j, a_k\right)\right|$ for reputation-seekers and wallflowers, respectively. Note that the utility function is unchanged otherwise; this implies that social information has no impact when decisions are not visible. We will discuss an alternative model where social information matters without visibility in the Extensions subsection below.

For reputation seekers, marginal reputation benefits remain the same as they were without social information: $r(a|s) = E(v|a) - E(v|a-1)$. On the other hand, marginal reputation benefits for reputation avoiders now depend heavily on social information. In particular $r(a|s) < 0$ for a that is higher than a_j and a_k. As a result, visibility will drive wallflowers to decrease their contribution when they are told that others gave little.

4.2.3 Testable prediction

Wallflower preferences imply that individuals avoid low or high contributions and seek out contributions within the range of the social information they observe. That is:

Prediction: Suppose person i is given information about person j and k's contributions (a_j and a_k where $a_k > a_j$) before choosing his own contribution a_i. Revealing a_i will increase $Pr(a_j < a_i < a_k)$.

Again, this prediction differs from what would be observed among reputation seekers. Since visibility leads to higher contributions for reputation-seeking types, revealing a_i will increase $Pr(a_i > a_k)$.

4.3 LABORATORY EXPERIMENT

We now turn to the design and results of our laboratory experiment. To test the implications of our model, it would be ideal to manipulate participants' beliefs about the average level of giving in a charitable giving task and test whether the impact of visibility is increasingly positive (negative) depending on how high (low) others' giving is. However, directly manipulating beliefs in the laboratory is not a straightforward task. We therefore assign subjects to groups and use the strategy method to elicit subjects' giving strategy *conditional on the contributions of their group members* who, in one of the treatments, will observe what they gave.[49] By doing so, we are able to observe what a particular subject would choose to do under a variety of social information contexts. In using the strategy method, we essentially test the impact of (hypothetical) social information at a within- rather than between-subject level.

[49] The conditional contribution elicitation is based on the design of Fischbacher et al. (2001), but with some important differences. Fischbacher et al. allow participants to condition on the mean of their group members' contributions whereas we allow participants to condition on every possible combination of group members' contributions.

4.3.1 Experimental Design

Subjects were recruited through the Pittsburgh Experimental Economics Laboratory (PEEL) database. In each session, fifteen subjects were seated at computer terminals upon arrival and randomly assigned to groups of three and identified only by anonymous subject IDs. We explained that the experiment would consist of a "giving task," during which they would have the opportunity to donate to a charitable cause, and a "guessing task," during which they would have the opportunity to earn "up to an additional $7."

The software then played a slideshow of a water project in Tingo Pucara, Ecuador, organized by Engineers Without Borders (EWB) Pittsburgh, our partner nonprofit for the laboratory experiment. After the slide show, we endowed participants with 10 one-dollar bills in an envelope, from which they would make their contributions. Any money that participants did not donate was theirs' to keep.

There are two treatments: *Control* and *Visibility*. The only difference between these two treatments is that *Visibility* treatment subjects know that their identity will be revealed to their group members and the experimenter after all decisions have been made.

In the control treatment, the experimenter read from the following script:

> *"At the end of today's session, you will leave your donation in its original envelope on your desk. The software will inform you of your group's total donation to Tingo Pucara."*

In the visibility treatment, the experimenter read from the following script:

> *"At the end of today's session, you and your group will sit down together around a table to submit your contributions. You will go to a different room with the experimenter who will then collect each group member's contributions, announce how much each person gave, and announce the total donation to Tingo Pucara. Your group members are the only participants who will observe how much you chose to give."*

After these preliminaries were completed, the decision-making portion of the experiment consisted of three phases: (1) unconditional contributions, (2) conditional contributions, and (3) belief elicitation. These decision tasks are explained in detail below. Participants did not learn of the details of any of these phases until they occurred nor were participants aware that the "giving task" would ultimately consist of both an unconditional and conditional contribution task.

We conducted five sessions of each of the two treatments with 15 participants per session. There were (roughly) equal proportions of males and females in all ten sessions, with 75 males and 75 females participating overall. All sessions were conducted in the Pittsburgh Experimental Economics Laboratory (PEEL) using z-Tree software. Subjects received a $5 show-up fee in addition to any money kept or earned during the experiment. Sessions lasted less than one hour.

In the *unconditional contribution* task, participants were simply asked to indicate how much of their $10 endowment they wanted to donate. We restricted contributions to multiples of 2; that is, participants could choose to give $0, $2, $4, $6, $8, or $10. This restriction and the small group sizes were chosen to limit the number of choices that participants would face in the conditional contribution task.

In the *conditional contribution* task, participants were given the opportunity to change their contribution based on what the other two members of their group chose. Participants choose contributions conditional on *every possible combination* of their group members' contributions from the unconditional phase – a total of 21 decisions. For instance, in the first screen (Figure 4.1), a subject is asked to assume that one of their group members gave $0 in the unconditional phase. A list of all possible unconditional contributions of the second member of their group ($0, $2, $4, $6, $8 and $10) is displayed, and the subject is asked to indicate her donation for each combination of hypothetical contributions. A similar series of screens then follows. In the second screen, subjects are asked to assume one group member gave $2 and to then indicate how she would contribute if the second group member gave $2, $4, $6, $8, or $10. The following screens present the rest of the scenarios, fixing one group member's contribution at $4, $6, $8 and $10.

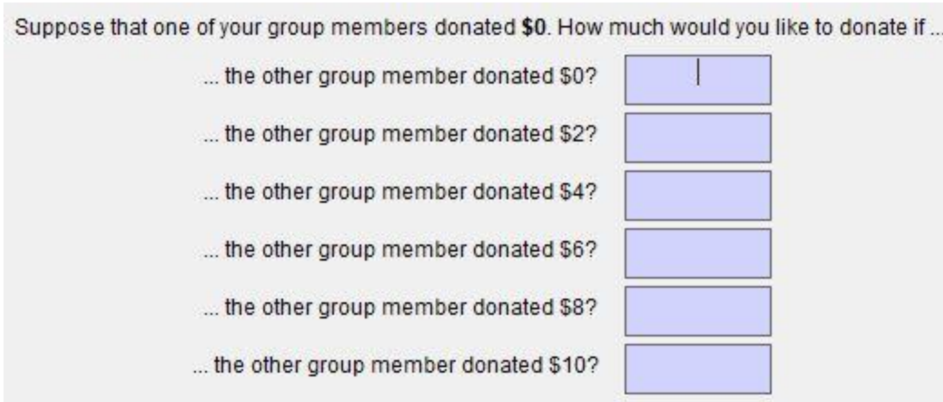

Figure 4.1. Conditional contribution entry

As in the conditional contribution design of Fischbacher et al. (2001), one member of each group was randomly selected at the end of the session to have her conditional contribution implemented. Thus, when participants submitted their contributions at the end of the experiment, two members of each group submitted the contribution they chose in the unconditional phase and the remaining member submitted the relevant conditional contribution. Even in the visibility treatment, participants never learned which member of their group was randomly selected to have his or her conditional contribution implemented.

In the *belief elicitation* task, participants were asked to guess the number of people who chose each of the possible unconditional contributions. Participants were informed that they had 14 tokens (one for each of the other participants in the room) to allocate across the possible unconditional contributions ($0, $2, $4, $6, $8, and $10). Each token that was placed correctly earned the participant $0.50. Denoting a participant's *reported guess* of the number of subjects who chose unconditional contribution k as g_k and the actual number of subjects as n_k, then the belief elicitation payoff can be expressed as:

$$\sum_k 0.5 * \min\{g_k, n_k\} \ for\ k \in \{0, 2, 4, 6, 8, 10\} \quad (7)$$

After completing the belief elicitation task, participants completed a brief survey where they indicated their gender and familiarity with the charitable cause.[50] They were next informed

[50] There is no evidence of a gender difference in familiarity to the organization (EWB). Nationally, 45% of EWB members are women.

of (1) their earnings from the belief elicitation task, (2) the actual contribution they would provide based on whether or not they were the randomly selected member of their group and, if so, the unconditional contributions of their group members, and (3) their group's total contribution to the cause. Donations were then collected according to the procedures described above.

4.3.2 Results

Our experiment is designed to examine how participants position themselves within the distribution of group members' contributions under visibility and not just whether visibility increases or decreases giving. Thus, our main analysis in this section centers on examining the relationship between participants' conditional contributions and the (hypothetical) choices of their partners. Despite this, it is instructive to first consider simple summary statistics and determinants of giving, after which we will turn to more specific hypotheses regarding participants' conditional contribution decisions. In some additional results, we also take advantage of the belief elicitation and examine the relationship between unconditional contributions and expectations about others' giving.

4.3.2.1 Summary statistics and determinants of giving

Table 4.1 reports summary statistics from all phases of the experiment: unconditional contributions, conditional contributions, and elicited beliefs. Recall that in the belief elicitation procedure, participants indicate their beliefs about how many of the 14 other participants in their session chose an unconditional contribution of $0, how many chose $2, etc. Because their responses across the possible contribution levels must add to 14, elicited beliefs can be interpreted as a participants' subjective probability distribution. Thus, we simply divide participants' reports by 14 to obtain participants' beliefs as probabilities. As a summary measure, we use these probabilities to construct each participant's expected value of others' unconditional contributions.

In both the unconditional and conditional phases of the experiment, the raw means of contributions are slightly higher under visibility but not significantly so. Similiarly, participants expect others to choose slightly (and insignificantly) higher contributions under visibility.

Table 4.1. Summary statistics

	Control	Visibility	p-value (two-tailed t-test)
Unconditional contribution	3.333	4.133	*0.16*
	(0.408)	(0.395)	
Conditional contribution[1]	3.010	3.641	*0.22*
	(0.381)	(0.348)	
Belief: E(uncond. cont.)	3.396	3.484	*0.79*
	(0.242)	(0.219)	
Observations	75	75	

Standard errors in parentheses.
[1] Reported standard errors are from the mean of within-subject averages of conditional contributions, where there is one observation per participant.

Of course, our experiment is designed to test how visibility interacts with social information. Thus, we turn to regression analysis in Table 4.2 to assess participants' reaction to social information and, in particular, how the impact of visibility varies with social information.

In all specifications of Table 4.2, we take "Conditional contribution" as the dependent variable and regress this on a dummy for being visible, the mean of group members' (hypothetical) contributions, and – in Columns 2 and 3 – the interaction thereof. Because each participant makes 21 conditional contribution decisions, we allow for clustered standard-errors at the individual level.

In Column 1, we simply regress Conditional Contributions on the visibility dummy and the group mean. As in the preceding means comparisons, this specification reveals that visibility has an insignificantly positive effect on giving. Column 1 also reveals a positive relationship between conditional contributions and the mean of group members' contributions.

In Column 2, we allow the impact of visibility to vary with social information. Upon doing so, we find that the impact of visibility depends on the mean of group members' contributions; as group members' contributions increase, visibility has an increasingly positive effect. More concretely: when the mean of group members' contributions is zero, visibility has a small, insignificantly negative effect on giving. On the other hand, when the mean of group members' contributions is ten, visibility increases contributions by $1.34 (p-value=0.017)[51].

Table 4.2. Determinants of conditional contributions (OLS)

VARIABLES	(1) Conditional contribution	(2) Conditional contribution	(3) Conditional contribution
Vis.	0.631	-0.0803	-0.0803
	(0.515)	(0.554)	(0.556)
Vis. X Group mean		0.142***	0.142***
		(0.0415)	(0.0417)
Group mean	0.0607***	-0.0105	
	(0.0216)	(0.0241)	
Constant	2.706***	3.062***	3.010***
	(0.391)	(0.390)	(0.381)
Decision FE's			X
Observations	3,150	3,150	3,150
R-squared	0.011	0.014	0.015

Standard errors in parentheses, clustered at individual-level.
*** p<0.01, ** p<0.05, * p<0.1

Column 3 include "Decision Fixed Effects" to assess the robustness of these results. Participants make 21 decisions; decision fixed effects are simply dummy variables for each of these decisions. This specification has the advantage of more flexibly accounting for social information. In particular, the Group Mean might be identical across two sets of decision (e.g., $0 and $10 versus $4 and $6), but the average response may differ. This would be captured by the decision fixed effects. Decision fixed effects prevent us from being able to identify the effect of the Group Mean in the Control treatment, but we can still observe the differential impact of the Group Mean conditional on being in the Visibility treatment. Results are similar to Column 2. Decision Fixed Effects are used in later sections of the paper.

In summary:

---

[51] This prediction is based on the linear combination of Vis+(VisXGroup Mean)X$10.

Lab Experiment Finding 1: Visibility does not lead to a general increase in contributions in this context. Instead, visibility only has a positive impact when participants expect group members to choose a high contribution.

This finding provides some evidence against simple reputation-seeking preferences, from which we might have expected a general increase in contributions with visibility. However, the positive relationship between social information and contributions documented in Table 4.2 (when visible) may stem from either wallflower preferences or from matching either of the group members' contributions (conformity). Alternatively, participants may hold more complex reputation-seeking preferences: systematically choosing a contribution just above the maximum of group members' contributions would also lead to the documented pattern. In the next subsection, we attempt to disentangle these explanations by more directly examining where participants position their contributions relative to group members.

4.3.2.2 Direct evidence of wallflower preferences

Our model suggests that wallflowers are drawn towards the action that would signal that the individual taking it is an "average type" relative to their group members. Reputation-seekers are drawn towards actions that signal high types. This implies our main test for the presence of wallflower preferences: given social information about player j and k's contributions (a_j, a_k), where $a_j \leq a_k$, visibility increases the probability that participant i chooses "middle" actions ($a_i$between a_j and a_k) and decreases the probability of extreme actions, that is, a_j that is larger than a_k .

We estimate a series of linear probability models assessing the likelihood that for some conditional decision a participant chooses a contribution within a particular range relative to her group members' (hypothetical) contributions. The main ranges of interest are: (1) less than the minimum ($a_i < a_j$), (2) equal to the minimum ($a_i = a_j$), (3) within range ($a_j < a_i < a_k$), (4) equal to the maximum ($a_i = a_j$), or (5) greater than the maximum ($a_i > a_k$). Thus, we estimate a set of regressions of the form:

$$y_{id}^{r} = \alpha + \beta_v(visibility_i) + \delta_d + \beta_b[beliefs_i] + \beta_f unfamiliar_i$$

where $y_{id}^{r} = 1$ if participant i chooses a contribution within the associated range r in conditional decision $d \in \{1,..,21\}$. Note that it is not always possible to choose a contribution within a particular range. For instance, when the minimum of partners' contributions is 0, it is impossible to choose a contribution "less than the minimum." These instances are excluded from the analysis.[52] As in Column 3 of Table 4.2, we include δ_d (decision fixed effects), controls for beliefs, and familiarity with the cause. We cluster standard errors at the individual level. The *Visibility* coefficient is of primary interest.

Panel A of Table 4.3 reports our main results. Consistent with our prediction, participants are significantly more likely to choose a contribution within the range of group members' contributions (Column 3). Visibility also moves subjects away from an extremely low (< Min.) contribution (Column 1). Contrary to reputation-seeking preferences, there is no evidence that visibility increases the likelihood of choosing a contribution that is high relative to group members (Column 5).

Panel B examines how people adjust to social information more directly and provides an opportunity to test our model in a more targeted manner. Because we can identify what an individual would choose in the absence of social information (unconditional contribution), we can identify the individuals who should be especially likely to move towards the center: the individuals whose unconditional contributions were just above ($2 higher than) the maximum of group member contributions and individuals whose unconditional contributions were just below ($2 below) the minimum. In Panel B, we modify the main specification (from Panel A) to allow for an interaction between visibility and dummy variables indicating that an individual's unconditional contribution is just below the minimum ("Just below") or just above the maximum

---

[52] Other exclusions: It is impossible to be "greater than maximum" when the maximum contribution is 10, so all such cases are excluded from "greater than max." regression. It is impossible to be strictly "within range" when the distance between partners' contributions is 0 or 2, so these cases are excluded from "within range" regression. Finally, there is no unique maximum or minimum when partners' contributions are identical, so these cases are excluded from the "equal to min." and "equal to max." regressions.

("Just above"). For brevity, in Panel B we simply report the marginal effects of visibility for each of these two groups.[53]

Table 4.3. Likelihood of choosing a contribution within a particular range

VARIABLES	(1) < Min.	(2) = Min.	(3) Within range	(4) = Max.	(5) > Max.
			PANEL A: Main specifications		
Vis.	-0.112*	0.0222	0.0720*	0.0258	-0.00889
	(0.0577)	(0.0233)	(0.0419)	(0.0248)	(0.0510)
Constant	0.611***	0.228***	0.244***	0.108***	0.212***
	(0.0420)	(0.0145)	(0.0289)	(0.0173)	(0.0368)
Observations	2,250	2,250	1,500	2,250	2,250
R-squared	0.178	0.050	0.141	0.036	0.107
			PANEL B: Assessing the impact of visibility on marginal choices		
Vis. X Just above	0.0195	0.0367	0.0241	0.194**	-0.266**
	(0.0667)	(0.0629)	(0.0784)	(0.0961)	(0.126)
Vis. X Just below	-0.128**	0.125**	-0.0309	0.00680	-0.00123
	(0.0572)	(0.0594)	(0.0313)	(0.00501)	(0.0131)
Constant	0.477***	0.279***	0.221***	0.110***	0.155***
	(0.0374)	(0.0318)	(0.0414)	(0.0250)	(0.0333)
Observations	1,500	900	450	900	1,500
R-squared	0.443	0.109	0.125	0.100	0.242

Standard errors in parentheses, clustered at individual-level.
Decision Fixed Effects and controls for beliefs and familiarity are included.
*** p<0.01, ** p<0.05, * p<0.1

Having identified the individuals who *should be impacted* by the interaction of visibility and social information, Panel B provides clear evidence of wallflower behavior. Individuals who would have otherwise chosen a contribution just below the minimum are (a) significantly *less* likely to stay below the minimum under visibility (Column 1) and (b) significantly *more* likely to move towards the center – choosing a contribution equal to the minimum (Column 2). This is consistent with wallflower behavior, but the decreased likelihood of choosing a low contribution is also consistent with reputation seeking preferences. A stronger test is provided by examining

[53] That is, the linear combinations "Vis + Vis X Max+2" and "Vis + Vis X Min-2".

individuals who, without social information, would have chosen a high (> max.) contribution. Walllflower types would be *less* likely to choose a contribution that exceeds the maximum after receiving social information. Indeed, Column 4 suggests that this is the case. Participants who would have given a high amount respond by decreasing their contribution and towards the center.

Alternatively, we can take individuals as the unit of observation and assess the fraction of contributions that a participant places within the range of partners' contributions. If wallflower preferences are prevalent, the average frequency of "within range" choices should increase with visibility. This is the case, as displayed in Panel A of 4.2. (Panels B and C are discussed in the next subsection.) The average participant in the visibility treatment places 32% of choices within range, while the average control treatment participant places only 24% of choices within range. This difference is significant at the 10% level.

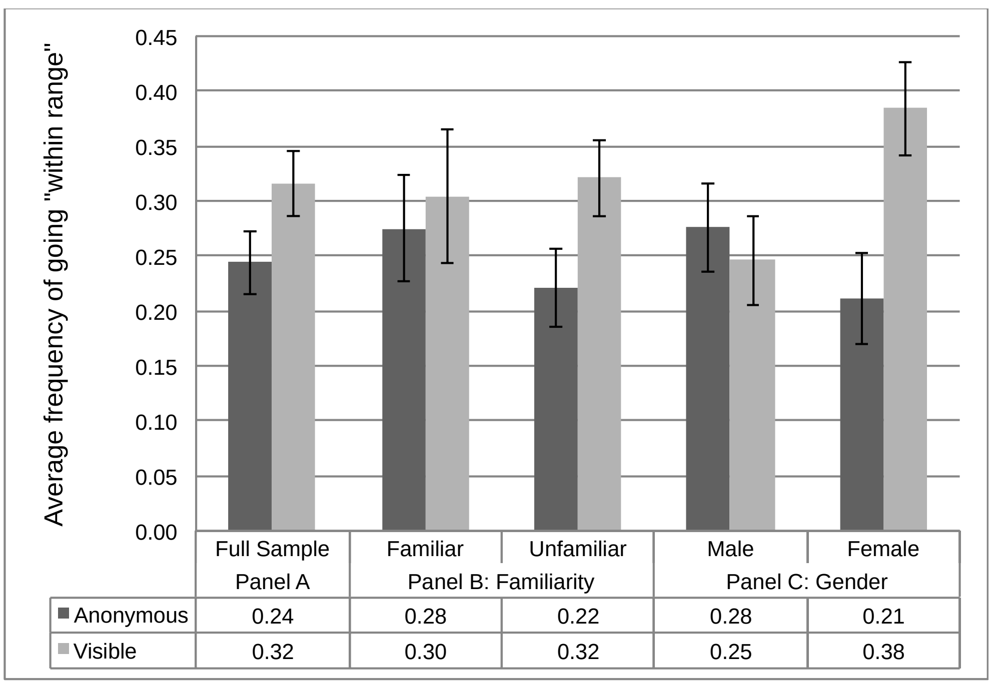

Figure 4.2. Individual level frequency of "within range" contributions

Lab Experiment Finding 2: Consistent with wallflower behavior (but not reputation seeking preferences), visibility induces participants to move towards the center of the distribution of group members' contributions. Participants are significantly more likely to choose a contribution strictly within the range of group members. Participants who were just above or just below group members' contributions are especially likely to move towards the center.

4.3.2.3 Heterogeneity in reputation concerns

The previous section provides evidence of the presence of wallflowers. In this subsection we ask: who are they? Is there heterogeneity in reputation concerns? Previous literature suggests two possible dimensions where this heterogeneity may be observed: task familiarity and gender. Existing work has shown that individuals who have less experience with a charitable cause are more likely to respond to social information about it (Shang & Croson, 2009). Moreover, if familiarity with the cause enhances intrinsic desire to give, higher familiarity may reduce the importance of reputation concerns in determining a donor's level of giving.

Research in psychology has suggested that women are more likely than men to conform when observed by a group (Eagly, 1978; Eagly & Carli, 1981; Bond & Smith, 1996). Croson and Gneezy (2009) review literature from psychology and economics showing that women's behavior shows higher sensitivity to changes in the social environment than men's.[54] It has also been shown that females tend to avoid competitive situations altogether when given the choice (Niederle & Vesterlund, 2007). Additionally, Kanthak and Woon (2012) show that females are reluctant to reveal their ability to compete in an election environment. If publication of prosocial behavior is perceived as a status competition, it is unsurprising to find that females "opt out" of this competition by choosing actions that are least likely to draw attention. The interaction of prosocial behavior and competition in altruism signaling can be seen in Pan and Houser (2011) laboratory public goods game where participants directly compete to be the "best giver". The authors find that when altruism signaling is possible, males (and not females) respond by

[54] Zetland and Della Giusta (2011) vary the salience of social information in a public goods game and show that only women change their behavior in response to this manipulation. Others have shown that females are significantly less likely to donate blood when a monetary payment is offered, while males are more likely to donate (Mellstrom and Johannesson (2008), Lacetera and Macis (2010)). This may be due to females internalizing social cues that blood donation should be given out of pure altruism and not for money. There are no observable gender differences in crowding out when extrinsic incentives are given in the form of a gift voucher (Lacetera and Macis (2010)) or when participants are allowed to donate the cash payment to charity (Mellstrom and Johannesson (2008)).

becoming more competitive and, as a result, provide higher contributions. Thus, existing literature suggests that women may be more likely than men to display wallflower preferences.

We test these hypotheses by modifying the main empirical specification from the Table 4.3 (Panel A) to allow for interactions between visibility and familiarity, and also visibility and gender. We also provide some additional evidence on gender differences in the unconditional contribution phase.

Results are presented in Table 4.4. As before, let (a_j, a_k) represent what person j and k hypothetically gave and assume $a_j \leq a_k$. We report only the marginal effects of visibility for each group (familiar and unfamiliar, male and female).

Table 4.4. Likelihood of choosing a contribution within a particular range **-** Heterogeneous responses

VARIABLES	(1) < Min.	(4) = Min.	(5) Within range	(6) = Max.	(3) > Max.
PANEL A: Differential response to visibility by Familiarity with Cause					
Vis. X Familiar	-0.134	0.0240	0.0293	0.0572	0.00815
	(0.0956)	(0.0430)	(0.0767)	(0.0433)	(0.0974)
Vis. X Unfamiliar	-0.131*	0.0137	0.100**	0.0191	0.00623
	(0.0688)	(0.0278)	(0.0496)	(0.0313)	(0.0563)
Constant	0.517***	0.202***	0.275***	0.131***	0.287***
	(0.0681)	(0.0221)	(0.0480)	(0.0219)	(0.0630)
Observations	2,250	2,250	1,500	2,250	2,250
R-squared	0.203	0.052	0.143	0.044	0.131
PANEL B: Differential response to visibility by Gender					
Vis. X Male	-0.0853	0.0371	-0.0304	0.0364	-0.00574
	(0.0778)	(0.0326)	(0.0571)	(0.0278)	(0.0700)
Vis. X Female	-0.136	0.00835	0.173***	0.0139	-0.0133
	(0.0839)	(0.0330)	(0.0589)	(0.0406)	(0.0742)
Constant	0.646***	0.239***	0.276***	0.0789***	0.188***
	(0.0532)	(0.0165)	(0.0403)	(0.0152)	(0.0458)
Observations	2,250	2,250	1,500	2,250	2,250
R-squared	0.188	0.053	0.155	0.041	0.110

Standard errors in parentheses, clustered at individual-level.
Decision Fixed Effects and controls for beliefs and familiarity are included.
*** p<0.01, ** p<0.05, * p<0.1

In Panel A, we interact "Vis." with a dummy variable indicating that the participant is unfamiliar with the cause ("Unfamiliar"). We find that individuals who are *unfamiliar* with the cause are less likely to give a low amount (Column 1) and more likely to choose a contribution within a_j and a_k (Column 3). This is not only consistent with results in the social information literature, but also suggests that unfamiliar individuals may be more affected by social information because their giving decision is more motivated by reputation concerns than those who familiar with the cause.

In Panel B, we add a control for "Female" and the interaction of Female and Visibility. We find that the impact of visibility reported in the previous subsection is almost entirely driven by women. Women are significantly more likely to choose a contribution within range when visible and (insignificantly) less likely to choose a contribution below the minimum or above the maximum of group members' contributions. While the estimates for men are imprecise and of much smaller magnitude, there is some evidence that men are more likely to match either the minimum or maximum of group members contributions.

These patterns can be seen in Panels B and C of Figure 4.2, which reports the average fraction of contributions that a participant places "within range" by treatment. Visibility significantly increases the fraction of contributions placed within range for women (p-val.=0.005) and participants unfamiliar with the cause (p-val.=0.049), but not men or individuals unfamiliar with the cause.

Lab Experiment Finding 3: Wallflower behavior is particularly prevalent amongst women and individuals unfamiliar with the cause. Both groups are significantly more likely to choose contributions within the range of group members' contributions when visible.

Women and individuals unfamiliar with the cause respond to visibility by moving towards the mean of group members' contributions. We now examine whether this heterogeneity in reputation concerns can also be detected in the absence of social information. To do so, we calculate "belief-normalized contributions": using elicited beliefs, we calculate each subjects' expectation of other's average unconditional contribution and subtract this quantity from subjects' *unconditional* giving. With wallflower preferences, visibility would compress belief-normalized contributions of wallflower types around 0; on the other hand, it would shift to belief-normalized contributions of reputation seeking types to the right. We find evidence of reputation heterogeneity in this context across gender but none across familiarity. Figure 4.3 illustrates that women's contributions are increasingly compressed around the center; this phenomenon is not true of men. Indeed, variance-ratio tests reveal that the variance of women's (but not men's) belief-normalized contributions decreases with visibility (one-tailed t-test p-value=0.061), while the mean of men's (but not women's) belief-normalized contributions increases (one-tailed t-test p-value=0.071).[55]

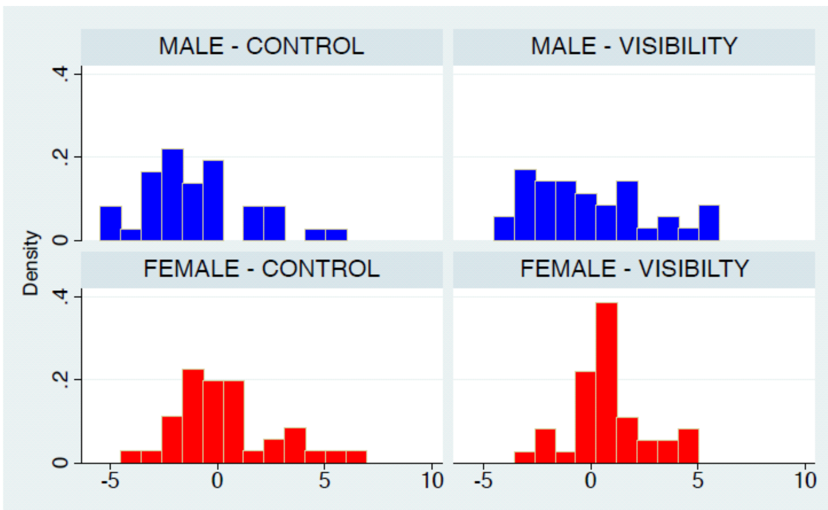

Figure 4.3. Distributions of "belief normalized contributions"

4.3.3 Summarizing lab results

In summary, our laboratory experiment provides evidence of wallflower behavior. Visibility does not always increase prosocial behavior. Instead, it induces participants to follow what they think others will do more closely. This conformity to social information is driven by participants' tendency to move towards "middle" actions when observed, suggesting a desire to minimize

[55] A Kolmogorov-Smirnov test rejects the hypothesis that women's "belief-normalized contributions" across control and visibility treatments are drawn from the same distribution (p-val.=0.65). This null hypothesis cannot be rejected for men (p-val.=0.475). We conducted similar tests for participants familiar and unfamiliar with the cause, but find no important differences. Visibility increases the mean of "belief-normalized" giving for both familiar and unfamiliar participants, but does not differentially impact the variance of giving across groups.

reputation signals. This is especially true of women and those who are unfamiliar with the nonprofit. There is additional evidence that women respond to visibility differently in the unconditional phase of the experiment. Visibility reduces the dispersion of women's contributions around their expectation of others' giving. Overall it appears that visibility amplifies the effect of social information – it lowers giving when others are contributing little and increases giving when others are contributing a lot.

4.4 GENERALIZING WALLFLOWER BEHAVIOR

In this section, we test the generalizability of the results from our laboratory experiment. We report the results of a small-scale fundraising experiment, using gender to identify individuals with wallflower preferences. (Information on familiarity with the cause is unavailable in our field experiment.)

We examine whether our findings on wallflower preferences translate to settings outside the lab in a small field experiment. Partnering with Engineers for a Sustainable World (ESW), we set up a computer game booth at a campus fair to raise money for a development project.[56] In the game, a 15x15 grid of jumbled letters hides 12 words. For every word found, 15 cents are donated to the project. When one word search puzzle is completed, another appears. Fair attendees can stop by the booth to play the game for as long as they wish. When participants click on a quit button, they are directed to a short demographic survey.

Participants are randomly assigned to one of two treatments: *Scores,* or *Names*.[57] In the Scores treatment, the $ amount raised by previous players is displayed before the game starts.[58] In the Names treatment, participants are first asked to enter their name; they are then shown the $

---

[56] The booth had five laptop computers programmed to display the word search game. Partitions are set up such that only one person can be at each computer.

[57] A *Baseline* treatment was also conducted, but is not discussed here. In the Baseline treatment participants simply played the wordsearch game without receiving any information about the donations of previous participants and without being asked to enter their name. In total we had 104 subjects, 71 in Scores and Names. 20 subjects without starting the survey and hence their gender information is not available.

[58] The listing for both *Scores* and *Names* includes only the last 10 people on that computer.

amount *and* names of previous players. They are aware that future participants will observe their name and donation.

On average, participants spend 5.5 minutes working and contribute $2.25, which corresponds to finding 15 words. A modal contribution emerges: roughly half of all participants stop at the end of the *first* puzzle, finding exactly 12 words and contributing $1.80. Male donation went from $2.79 in Scores (N=17) to $2.94 in Names (N=10). However, female donation decreased from $3.15 in Scores (N=12) to $2.08 in Names (N=12). Figure 4.4 shows that the drop in giving is not due to a simple downward shift: when names are displayed, females in the Names treatment are significantly more likely to choose the modal contribution of $1.80. The same is not true for male.[59]

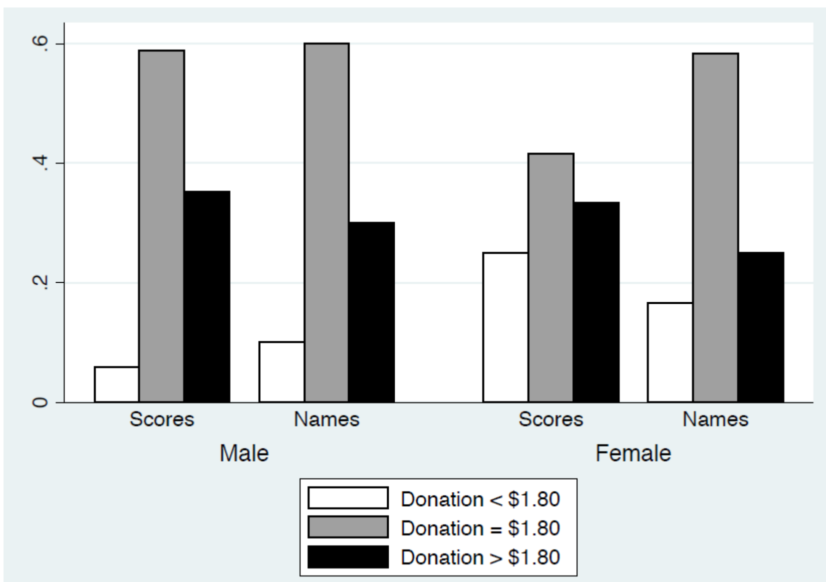

Figure 4.4. Freq. of donations below, equal to, and above $1.80 by treatment and gender

[59] Though not displayed here, the significant increase in women's likelihood of choosing $1.80 in response to the names treatment is demonstrated through a probit estimation, which includes computer fixed effects and a control for the number of other individuals in the booth at the same time. The marginal impact of the Names treatment on women's likelihood of choosing $1.80 is significant at the 10%.

4.5 CONCLUSION

In this paper, we revisit the question of how visibility impacts prosocial behavior. We theoretically propose a novel response to visibility: reputation avoidance, or "wallflower" behavior. We suggest that some individuals are averse to standing out in either a negative way (appearing selfish) or positive way ("showing off") when visible. These preferences imply that individuals will respond to visibility by avoiding actions that are either very low or very high relative to others, thereby pushing individuals' choices towards the center of the distribution of actions (or what they expect to be the center.)

We provide evidence of such behavior in a laboratory experiment. Participants are able to choose a donation to a charitable cause conditional on every possible combination of the donations of their two group members, who – in the Visibility treatment – will observe how much they donated. Consistent with the theoretical predictions, we find that visibility significantly increases the likelihood that participants choose a contribution within the range of group members' contributions, thereby avoiding contributions that could be perceived as very high or very low.

In additional analyses, we find some evidence of heterogeneity in reputation avoidance. Individuals who were unfamiliar with the charitable cause in our laboratory experiment are more likely to display wallflower behavior. We have stronger evidence on heterogeneity across genders. Women in particular consistently respond to visibility by choosing a contribution strictly within the range of group members' contributions. We have less conclusive evidence on the behavior of men, but they appear to be more likely to choose a contribution at the boundaries of the range of group members' contributions (either matching the minimum or maximum). There are also gender differences in contribution decisions when participants do not know how much their partners have given (the unconditional decisions). We observe participants' beliefs about partners' giving and find that, with visibility, the distribution of women's contributions is more compressed around their expectation of the average level of giving.

The evidence we find of reputation avoidance may help explain the divergent results in the literature. While most of the existing experimental work tests the hypothesis that visibility increases giving, we suggest that this may not be the right question to ask. Differences in the impact of visibility in the public goods game and dictator game, for instance, may be explained by differences in participants' beliefs about the average level of giving across the two environments. According to the wallflower model, the impact of visibility depends on whether individuals perceive the average level of giving as high or low. As a result, visibility may have positive, negative, or no impact, depending on the environment.

We find evidence of this implication in our data. In the conditional contribution decisions, visibility only has a positive impact when the mean of partners' giving is high. We also see evidence of this implication in our field experiment, particularly for women. Women's donations decrease because they are drawn to the relatively low modal contribution that emerges.

These findings are far from being of purely theoretical interest. Outside of the experimental laboratory, prosocial behavior is rarely anonymous. Fundraisers often purposefully increase the visibility of contributions by publishing donor lists for instance. Other activities – such as volunteering or going to a polling location to vote, which are more often than not carried out in public and in the presence of peers – are visible by their very nature. Thus, a deeper understanding of the impact of such visibility is critical.

Our findings suggest that care must be taken in manipulating the degree of visibility of prosocial behavior. This is particularly true in prosocial environments where females are overrepresented, as is the case in the nonprofit sector. Indeed, as Gibelman (2000) demonstrates, though females make up a large majority of the nonprofit sector, upper management positions continue to be male-dominated. While we certainly do not claim to fully explain this phenomenon, our finding that females prefer to avoid drawing attention to themselves when prosocial behavior is visible, may play an important role. In the context of charitable giving, males and females have different preferences in the charitable causes they support (Andreoni et al., 2003). This fact, combined with our results, suggests that causes that attract more female givers may require a very different fundraising strategy than those that attract males.

APPENDIX

ADDITIONAL RESULTS FROM CHAPTER 2

Table A1: Main government outcomes – additional time varying controls

	(1)	(2)	(3)
VARIABLES	Revenue	Educ. exp.	Non-educ. exp.
Edulot	0.0133	0.00215	0.0143
	(0.0178)	(0.0324)	(0.0200)
Pre-treat trend	-0.00120	0.00273	-0.00574**
(Years before Edulot)	(0.00253)	(0.00202)	(0.00260)
Post-treat trend	0.00556***	0.00714	0.00348
(Years after Edulot)	(0.00175)	(0.00545)	(0.00288)
Non-educ. lottery	-0.00533	-0.0272	0.0202
	(0.0255)	(0.0439)	(0.0269)
Population	-0.127	-0.495***	0.163
	(0.106)	(0.156)	(0.178)
Income	0.941***	0.738***	0.510**
	(0.215)	(0.274)	(0.213)
Other revenue		0.203***	0.224***
		(0.0675)	(0.0596)
Total students	-0.247**	0.288*	-0.326**
	(0.0927)	(0.148)	(0.125)
Unemp. rate	0.00109	-0.0379	0.0488**
	(0.0336)	(0.0333)	(0.0240)
Republican gov.	-0.0140*	-0.00659	-0.0224**
	(0.00734)	(0.0135)	(0.00869)
Constant	3.417***	1.106	0.572
	(1.169)	(1.938)	(2.466)
Observations	950	950	950
R-squared	0.963	0.957	0.980

Robust standard errors (clustered at the state-level) in parentheses.
All continous controls and outcome variables are in logs. All monetary variables (including outcome variables) are measured at the per-capita level.
*** p<0.01, ** p<0.05, * p<0.1

Table A2: Education expenditures in an earlier time period for comparison with existing literature

	Years: 1989-2008 (as covered in this paper)			Years: 1976-2000 (as covered in Novarro and Evans & Zhang)		
	(1)	(2)	(3)	(4)	(5)	(6)
	Educ. exp.	Elem. educ. exp.	Higher educ. exp.	Educ. exp.	Elem. educ. exp.	Higher educ. exp.
Edulot	-0.00692	-0.0574	0.0208	0.00811	0.378*	0.0130
	(0.0314)	(0.0578)	(0.0355)	(0.0298)	(0.199)	(0.0307)
Obs.	950	950	950	1,250	1,092	1,250
R-squared	0.952	0.882	0.966	0.978	0.900	0.972

All specifications include the same controls and fixed effects as those reported in Table 2.
Education expenditure variables are in logs and measured at per capita level.
Robust standard errors (clustered at state-level) in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table A3: The impact of an education lottery on contributions received -- Allowing for treatment effects 2 years before and after the treatment date

VARIABLES	(1) Contributions received: Educ. orgs.	(2) Contributions received: Educ. orgs.
Edulot – 2 years		0.00414
		(0.0328)
Edulot	-0.0689**	-0.0475***
	(0.0279)	(0.0168)
Edulot + 2 years		-0.0451
		(0.0333)
Observations	27,905	27,905
R-squared	0.155	0.155

Empirical specifications match those reported in Table 10 (Column 2): IV-fixed effects regressions with fixed effects at the nonprofit level. The dependent variable is donations received by education related organizations. Column 1 matches the main result reported in Table 10. Column 2 extends this specification to allow for differential pre- and post-treatment trends.

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A4: Contributions received as a function of education-earmarked lottery proceeds

	(1)	(2)
VARIABLES	Contributions received: Educ. orgs.	Contributions received: Non-educ. orgs.
	Panel A: Fixed effects regressions	
ln(Edulot proceeds)	-0.00632***	0.000901
	(0.00204)	(0.00206)
	Panel B: IV Fixed effects regressions	
ln(Edulot proceeds)	-0.00525**	0.00119
	(0.00216)	(0.00307)

Unit of observation: nonprofit firm. Dependent variable: donations received. Empirical specifications match those reported in the text either without accounting for fundraising (Panel A) or accounting for fundraising through the IV strategy (Panel B).

Robust standard errors (clustered at state-level) in parentheses

*** p<0.01, ** p<0.05, * p<0.1

BIBLIOGRAPHY

J. Alm, M. McKee, and M. Skidmore. Fiscal pressure, tax competition, and the introduction of state lotteries. *National Tax Journal*, 46:463--463, 1993.

F. Alpizar, F. Carlsson, and O. Johansson-Stenman. Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics*, 92(5-6):1047--1060, 2008.

J. Andreoni. Giving with impure altruism: applications to charity and ricardian equivalence. *The Journal of Political Economy*, page 1447--1458, 1989.

J. Andreoni. Philanthropy. *Handbook on the Economics of Giving, Reciprocity and Altruism*, 2:1201--1269, 2006.

J. Andreoni, E. Brown, and I. Rischall. Charitable giving by married couples: who decides and why does it matter?. *Journal of Human Resources*, 38(1):111, 2003.

J. Andreoni, A. Payne, and S. Smith. Do grants to charities crowd out other income? evidence from the uk. *CMPO Working Paper*, (13/301)2013.

J. Andreoni and A.A. Payne. Do government grants to private charities crowd out giving or fund-raising?. *The American Economic Review*, 93(3):792--812, 2003.

J. Andreoni and A.A. Payne. Is crowding out due entirely to fundraising? Evidence from a panel of charities. *Journal of Public Economics*, 2011.

J. Andreoni and R. Petrie. Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of Public Economics*, 88(7):1605--1623, 2004.

R. Benabou and J. Tirole. Incentives and prosocial behavior. *The American economic review*, 96(5):1652--1678, 2006.

M. Benz. Not for the profit, but for the satisfaction?--evidence on worker well-being in non-profit firms. *Kyklos*, 58(2):155--176, 2005.

T. Bergstrom, L. Blume, and H. Varian. On the private provision of public goods. *Journal of public economics*, 29(1):25--49, 1986.

M.O. Borg, P.M. Mason, and S.L. Shapiro. *The economic consequences of state lotteries*. Greenwood Publishing Group, 1991.

G.J. Borjas, HE Frech III, and P.B. Ginsburg. Property rights and wages: the case of nursing homes. *Journal of Human Resources*, page 231--246, 1983.

C.T. Clotfelter and P.J. Cook. On the economics of state lotteries. *The Journal of Economic Perspectives*, 4(4):105--119, 1990.

C.T. Clotfelter and P.J. Cook. *Selling hope: state lotteries in america*. Harvard Univ Press, 1991.

C.C. Coughlin, T.A. Garrett, and R. Hernández-Murillo. The geography, economics, and politics of lottery adoption. *REVIEW-FEDERAL RESERVE BANK OF SAINT LOUIS*, 88(3):165, 2006.

R. Croson and U. Gneezy. Gender differences in preferences. *Journal of Economic Literature*,

47(2):448--474, 2009.
J. Delfgaauw and R. Dur. Signaling and screening of workers' motivation. *Journal of Economic Behavior & Organization*, 62(4):605--624, 2007.
M. Dufwenberg and A. Muren. Generosity, anonymity, gender. *Journal of Economic Behavior & Organization*, 61(1):42--49, 2006.
O.H. Erekson, K.M. DeShano, G. Platt, and A.L. Ziegert. Fungibility of lottery revenues and support of public education. *Journal of Education Finance*, page 301--311, 2002.
O.H. Erekson, K.M. DeShano, G. Platt, and A.L. Ziegert. Fungibility of lottery revenues and support of public education. *Journal of Education Finance*, page 301--311, 2002.
W.N. Evans and P. Zhang. The impact of earmarked lottery revenue on k-12 educational expenditures. *Education Finance and Policy*, 2(1):40--73, 2007.
U. Fischbacher, S. Gachter, and E. Fehr. Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71(3):397--404, 2001.
R.H. Frank. What price the moral high ground?. *Southern Economic Journal*, 63(1)1996.
S. Gachter and E. Fehr. Collective action as a social exchange. *Journal of Economic Behavior & Organization*, 39(4):341--369, 1999.
T.A. Garrett. Earmarked lottery revenues for education: a new test of fungibility. *Journal of Education Finance*, page 219--238, 2001.
M. Gibelman. The Nonprofit Sector and Gender Discrimination. *Nonprofit Management and Leadership*, 10(3):251--269, 2000.
J.H. Goddeeris. Compensating differentials and self-selection: An application to lawyers. *The Journal of Political Economy*, 96(2):411--428, 1988.
D.C. Grabowski and R.A. Hirth. Competitive spillovers across non-profit and for-profit nursing homes. *Journal of Health Economics*, 22(1):1--22, 2003.
P. Gregg, P. Grout, A. Ratcliffe, S. Smith, and F. Windmeijer. How important is pro-social behaviour in the delivery of public services?. *Journal of public economics*, 2011.
G. Heutel. Crowding out and crowding in of private donations and government grants. Technical report, National Bureau of Economic Research, 2009.
A. Heyes. The economics of vocation or why is a badly paid nurse a good nurse?. *Journal of Health Economics*, 24(3):561--569, 2005.
A.G. Holtmann and T.L. Idson. Wage determination of registered nurses in proprietary and nonprofit nursing homes. *The Journal of Human Resources*, 28(1):55--79, 1993.
A.G. Holtmann and T.L. Idson. Wage determination of registered nurses in proprietary and nonprofit nursing homes. *Journal of Human Resources*, page 55--79, 1993.
D.J. Houston. ``Walking the walk" of public service motivation: public employees and charitable gifts of time, blood, and money. *Journal of Public Administration Research and Theory*, 16(1):67--86, 2006.
Craig E Landry and Michael K Price. Earmarking lottery proceeds for public goods: empirical evidence from us lotto expenditures. *Economics Letters*, 95(3):451--455, 2007.
J. Lanfranchi, M. Narcy, and M. Larguem. Shedding new light on intrinsic motivation to work: evidence from a discrete choice experiment. *Kyklos*, 63(1):75--93, 2010.
A. Lange, J.A. List, and M.K. Price. Using lotteries to finance public goods: theory and experimental evidence. *International Economic Review*, 48(3):901--927, 2007.
L. Leete. Whither the nonprofit wage differential? Estimates from the 1990 census. *Journal of Labor Economics*, page 136--170, 2001.
E.S. Lin and S.Y. Wu. Lottery expenses and charitable contributions--taiwan's experience.

Applied Economics, 39(17):2241--2251, 2007.
S. Linardi and M.A. McConnell. No excuses for good behavior: volunteering and the social environment. *Journal of Public Economics*, 95(5-6):445--454, 2011.
H. Monti. Environmental policy and giving: does government spending affect charitable donations?. , 2010.
J. Morgan. Financing public goods by means of lotteries. *Review of Economic Studies*, 67(4):761--784, 2000.
J. Morgan and M. Sefton. Funding public goods with lotteries: experimental evidence. *The Review of Economic Studies*, 67(4):785, 2000.
M. Narcy. Would nonprofit workers accept to earn less? evidence from france. *Applied Economics*, 43(3):313--326, 2011.
M. Niederle and L. Vesterlund. Do Women Shy Away from Competition? Do Men Compete Too Much?*. *The Quarterly Journal of Economics*, 122(3):1067--1101, 2007.
C. Noussair and S. Tucker. Public observability of decisions and voluntary contributions in a multiperiod context. *Public Finance Review*, 35(2):176--198, 2007.
N.K. Novarro. Earmarked lottery profits: a good bet for education finance?. *Journal of Education Finance*, page 23--44, 2005.
X.S. Pan and D. Houser. Competition for trophies triggers male generosity. *PloS one*, 6(4):e18050, 2011.
A.E. Preston. The effects of property rights on labor costs of nonprofit firms: An application to the day care industry. *The Journal of Industrial Economics*, 36(3):337--350, 1988.
A.E. Preston. The nonprofit worker in a for-profit world. *Journal of Labor Economics*, 7(4):438--463, 1989.
M. Rege and K. Telle. The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics*, 88(7-8):1625--1644, 2004.
C.J. Ruhm and C. Borkoski. Compensation in the nonprofit sector. *Journal of Human Resources*, 38(4):992, 2003.
D. Serra, P. Serneels, and A. Barr. Intrinsic motivations and the non-profit health sector: evidence from ethiopia. *Personality and Individual Differences*, 51(3):309--314, 2011.
L. Shi. Monetary rewards, image concern, and intrinsic motivation: evidence from a survey on blood donation. *Working Papers*, 2011.
F.A. Sloan, G.A. Picone, D.H. Taylor Jr, and S.Y. Chou. Hospital ownership and cost and quality of care: is there a dime's worth of difference?. *Journal of health economics*, 20(1):1--21, 2001.
A.R. Soetevent. Anonymity in giving in a natural context--a field experiment in 30 churches. *Journal of Public Economics*, 89(11-12):2301--2323, 2005.
C.J. Spindler. The lottery and education: robbing peter to pay paul?. *Public Budgeting & Finance*, 15(3):54--62, 1995.
C.M. Tolbert, M. Sizer, and United States. Dept. of Agriculture. Economic Research Service. Rural Economy Division. *Us commuting zones and labor market areas: a 1990 update*. Economic Research Service, Rural Economy Division, 1996.
K.R. Troske. Evidence on the employer size-wage premium from worker-establishment matched data. *Review of Economics and Statistics*, 81(1):15--26, 1999.
B.A. Weisbrod. Nonprofit and proprietary sector behavior: Wage differentials among lawyers. *Journal of Labor Economics*, 1(3):246--263, 1983.
S.Y. Wu. Does charitable gambling crowd out charitable donations? Using matching to analyze a

natural experiment. , 2012.