ESSAYS ON THE ECONOMICS OF GENDER AND PARENTING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

BY

JORGEN MICHAEL HARRIS

AUGUST 2019

ESSAYS ON THE ECONOMICS OF GENDER AND PARENTING

Jorgen Michael Harris, Ph.D. 2019

In my first chapter, I present the first causal evidence on the effect of the entry of women into occupations on the wages of those occupations. In particular, I examine the effect of changes in the gender composition of an occupation on wages for men and women within the occupation. To determine the causal effect of a change in gender composition, I construct a shift-share instrument by using the dramatic increase in the relative educational attainment and workforce participation of women from 1960-2010 to instrument for changes in the gender composition of occupations with different levels of "exposure" to increased female work and education. I find evidence that a 10 percentage-point increase in the fraction of females within an occupation leads to an 8 percent decrease in average male wage and a 6 percent decrease in average female wage in the concurrent census year. Over the 10 years following the change in the gender composition, I find that the effect of such an increase in the fraction of females persists for male workers and grows for female workers, leading to an 8 percent decrease in male wages and an 11 percent decrease in female wages. I present suggestive evidence attributing this finding to effects of gender composition on the prestige and amenity value of occupations.

In my second chapter, (co-authored with Eleonora Patacchini and Marco Battaglini) we study the effect of hearing cases alongside female judicial colleagues on the probability that a Federal judge hires a female law clerk. Federal judges are assigned to cases and to judicial panels at random and have few limitations on their choices of law clerks: these two features make the Federal court system a unique environment in which to study the effect of professional interactions and beliefs in organizations. For our analysis, we constructed a unique dataset by aggregating federal case records from

2007-2017 to collect information on federal judicial panels, and by merging this data with judicial hiring information from the Judicial Yellow Books, a directory of federal judges and clerks. We find that a one standard deviation increase in the fraction of co-panelists who are female increases a judge's likelihood of hiring a female clerk by 4 percentage points.

In my third chapter, I develop and test a model that explains differences in parenting style by socioeconomic status. Spanking, severe discipline, and other forms of "Authoritarian Parenting" are more common among low income and black parents than among high income and white and Hispanic parents. Although they are associated with increased short-term obedience, these "Authoritarian" parenting strategies are also associated with lower levels of cognitive development, self-esteem, school performance and "moral internalization." I construct a multi-stage parenting model in which children choose levels of school effort and delinquent behavior while heavily discounting future consequences, and parents altruistically regulate their children both by disciplining them and investing in their self-control. The model predicts that parents employ more discipline when the negative effects of child delinquent behavior are large. I test this model by measuring the effect of school safety (which influences the cost of child delinquent behavior) on parenting practices, measured in the Los Angeles Family and Neighborhood Survey (LA.FANS). Because school safety affects parents only through its effect on children, the relationship between school safety and parenting style should not reflect parents' stress, social isolation, or economic circumstances. Controlling for family, neighborhood and school characteristics, I find that a 1 standard deviation increase in school disorder is associated with a 0.11 standard deviation increase in harsh parental discipline, with larger effects for poor and Black households. I argue that this effect is driven by parents' concern about the cost of their child's misbehavior.

BIOGRAPHICAL SKETCH

Jorgen Michael Harris is a Ph.D. Candidate in Economics at Cornell University. Prior to beginning his Ph.D. at Cornell, Jorgen received an A.B. in Economics at the University of Chicago. He has worked at RCF Economic and Financial Consulting, a Chicago-based organization with an emphasis on public and environmental policy issues, at MDRC, a nonprofit organization that evaluates public policies aimed at addressing poverty and inequality, and the US Census Bureau.

Dedicated to Julio Iglesias, Souvik (Ja-O) Basu, Lucas Whaley, Daniel Karlin, Louis Kelly, Reginald Bell, Conrad (Focus) Marhone, Cody Testerman, and Lucious Peters.

ACKNOWLEDGEMENTS

The work contained in this thesis would not have been possible without the support of a tremendous number of generous, thoughtful, kind, and insightful people. Everyone mentioned here, as well as many others, has made a profound impact on my life and my education, for which I am deeply grateful.

First, I have been incredibly fortunate to have the support of a wonderful thesis committee, consisting of Michael Lovenheim, Francine Blau, and Mallika Thomas. Mike, you have shown more faith in me than I've ever felt warranted. As long as I've known you, you have treated me like my ideas were valuable and my work was promising, even when there was very little empirical evidence to support doing so. I have felt safe coming to you with every kind of problem, and I do not know how I could have made it through graduate school without your support. Fran, you have been an incredible role model as a scientist and as a human being. Every time I have gone into your office, I have left with a deeper understanding of what I am doing, a greater awareness of everything I don't know, and an awe at what real expertise looks like. Mallika, many of my favorite moments in grad school have come during conversations in your office that have veered wildly away from their initial purpose toward a seminar, a paper, or an idea. The chance to watch you think through problems has been priceless, and your example has made me a better researcher and a better teacher. I am vastly richer for having known all three of you.

In addition, many other wonderful current and former members of the Cornell faculty have generously given time and attention to my work and my career path. John Abowd and Lars Vilhuber showed me what it looks like to pursue a meaningful agenda, taught me how to manage a project and write organized code, and worked through my earliest and worst ideas with me. Evan Riehl, Richard Mansfield, and Victoria Prowse

gave incredible amounts of their time and energy to support my development and the development of other labor students, both through the labor work in progress seminar and through many conversations about my work. Eleonora Patacchini and Marco Battaglini have been fantastic coauthors, and have taught me an incredible amount about how to get good research done. Douglas Miller, Christopher Barrett, and Steve Coate have given tons of time to helping make my work better. And Kaushik Basu has listened to everything from my half-baked ideas to my un-baked ideas with patience, interest, and kindness.

Cornell's graduate student community has been a consistent source of support, ideas, and inspiration. I have learned as much from you as from anyone. I want to especially thank my wonderful office mates, Caroline Walker, Angela Cools, and Miriam Larson-Koester—thank you for letting me subject you to every dumb idea that popped into my head, for giving encouragement and support whenever I needed it, and for giving me a reason to be excited to come to work. Vesall Nourani—thank you for letting me feed off of your depth of thought and your clarity of purpose (and off your grocery bill). Abhishek Ananth—I am in awe of your deep understanding of economic theory, and have secretly liked all of your jokes. Fikri Pitsuwan, Adam Shumway, Nikolaj Boboshko Amani Moin, and Isha Agrawal—thank you for listening to my ideas, sharing yours with me, and being great people. David Jaume—your insight, depth, and ambition of thought have kept me focused on doing work that matters. Tamara McGavock—you were the perfect mentor when I was starting my Ph.D., and have remained my role model for how to be a good scholar and a good person. Rhiannon Jerch—I cannot express how much it's meant to me to have you as a sounding board and as a friend.

To Ramya Sampath, Alexander El Nabli, and Andrew Alexander, thank you for being there for me when I most needed a friend. I love you guys.

I had the help of a fantastic team of undergraduate research assistants on the second chapter of this dissertation. Sam Markiewitz, Becky Cardinali, Cenjia Lu, and Ignasi Merediz Sola: thank you for your hard work, creativity, and patience.

I also want to thank a few of the people who got me started on this path before I came to graduate school. George Tolley, Sabina Shaikh, John List, Karna Basu, Victor Lima, Charles Michalopolous, and Roland Lewin—I never would have started down this path without your guidance, mentorship and inspiration.

To my parents, Jeff and Julie Harris, and my sister, Jennica Harris, thank you so much for your love, support, and guidance. You make me better all the time.

Finally[1], to my wife, Anna Sarfaty, every part of my life is richer because of you. I am better because of you. You have given me the confidence and encouragement to persevere through the hardest parts of the process, you have helped me be true to myself in what I pursue and how I pursue it, you have listened to my whining and complaining with grace, and you have helped me maintain perspective. You are my best friend and I love you.

[1] I also want to thank my two wonderful dogs, Grilka and Vashti, but doing so feels a bit unprofessional so I'm hiding it in a footnote. They loved me just as much when my work was going badly as they did when it was going well.

ACKNOWLEDGEMENT OF DATA AND FUNDING SOURCES

My research has been generously supported by the L.R. "Red" Wilson M.A. '67 Excellence in Economics Medal. This support has been extremely helpful, and has supported my purchase of equipment and data, as well as travel to conferences.

The second chapter of this dissertation was supported by a grant from the Cornell Institute for Social Sciences small grant competition. This support was essential to the completion of this work, and supported the purchase of data and the funding of an excellent team of undergraduate Research Assistants.

The third chapter of this dissertation uses data collected from the Los Angeles Family and Neighborhood Survey (LA.FANS). Data collection for L.A.FANS was funded primarily by grants HD35944 and HD49865 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, grant AG22005 from the National Institute on Aging, and grant ES13907 from the National Institute of Environmental Health Science. Information on how to obtain the L.A.FANS data files is available on the L.A.FANS website(lasurvey.rand.org).

Contents

Chapter 1: Do Wages Fall When Women Enter an Occupation?

ABSTRACT

I present the first causal evidence on the effect of the entry of women into occupations on the wages of those occupations. In particular, I examine the effect of changes in the gender composition of an occupation on wages for men and women within the occupation. To determine the causal effect of a change in gender composition, I construct a shift-share instrument by using the dramatic increase in the relative educational attainment and workforce participation of women from 1960-2010 to instrument for changes in the gender composition of occupations with different levels of "exposure" to increased female work and education. I find evidence that a 10 percentage-point increase in the fraction of females within an occupation leads to an 8 percent decrease in average male wage and a 6 percent decrease in average female wage in the concurrent census year. Over the 10 years following the change in the gender composition, I find that the effect of such an increase in the fraction of females persists for male workers and grows for female workers, leading to an 8 percent decrease in male wages and an 11 percent decrease in female wages. I present suggestive evidence attributing this finding to effects of gender composition on the prestige and amenity value of occupations.

JEL Codes: J16, J24, J31, J71, J82, J110

Keywords: Economics of Gender, Occupational Choice, Human Capital, Wage Differentials, Discrimination, Labor Force Composition

Section 1.1: Introduction

Among full-time workers, women earn 19% less than men on average (Blau & Kahn, 2016). While this wage gap has closed since the 1960s, it has changed less than would be expected based on women's relative gains in observed characteristics that are related to wage, particularly education and experience. One of the primary contributors to the remaining gender wage gap is the fact that women work in lower-paying occupations and industries than do men, with occupation and industry accounting for 51% of the remaining wage gap (Blau & Kahn, 2016). As a result, understanding why women work in lower-paying occupations than men is crucial to understanding why women earn less than men.

There are two broad explanations for the negative relationship between fraction female and pay in occupations. One explanation is that differences in the occupations of men and women reflect differences in the constraints, preferences and abilities of male and female workers. In this case, women's greater household responsibilities would lead female workers to work fewer hours, expend less effort at work, and to choose less demanding, lower productivity and lower paying jobs relative to men with equal levels of human capital (Becker, 1985).[2] Alternately, these occupation choices may reflect discrimination. In particular, women may face greater barriers in high-paying male-dominated occupations than they do in similarly skillful but lower-paying female occupations. This can take various forms, such as barriers to promotion (Maume, 1999), reduced access to mentorship and networking (Chen, et al., 2015), or higher rates of discrimination and sexual harassment (Parker, 2018).

[2] Relative to men, women are more likely to choose occupations that allow career interruptions (Görlich & De Grip, 2007), do not require long or inflexible working hours (Goldin C. , 2014), and do not require intense competition or uncertain rewards (Niederle & Vesterlund, 2005), leading to lower wages.

Both of these views assume that the characteristics of occupations are unaffected by their gender composition. However, the presence of women in an occupation may change that occupation's characteristics, including pay. In particular, the presence of women may lower the occupation's perceived importance and difficulty, and/or alter working conditions to reflect the preferences of female workers. If the presence of women in an occupation causes a decline in that occupation's wage, the returns to preferences and skills that lead workers to choose female-dominated occupations would fall, exacerbating the gender wage gap. However, because this process would not generate a premium for workers on the margin between female-dominated and male-dominated occupations, analyses that examine the wages of similar individuals in different occupations will attribute these wage differences to the characteristics of individuals. While a large sociology literature shows a negative correlation between changes in average male and female wage and changes in the female share of an occupation's workforce, no previous work has estimated this relationship using plausibly exogenous variation in the gender composition of occupations.[3]

I present evidence on the effect of changes in gender composition on the average wage of an occupation that more plausibly handles the selection problems that have plagued earlier research. I estimate a panel regression of the log average male and female wage of older workers in an occupation on the fraction of female workers among younger workers in that occupation[4]. Because changes in the gender composition of an

[3] Previous studies have either depended on cross-sectional variation in gender composition and wage (Cohen & Huffman, 2003), (England P. , 1992), on time-series variation (Catanzarite, 2003), (England, Allison, & Wu, 2007), (Karlin, England, & Richardson, 2002), or on panel data (Levanon & Allison, 2009). (Tam, 1997) shows that these methods are sensitive to changes in the set of included covariates.

[4] I separate older and younger workers in order to account for the possibility that the characteristics of the male and female labor force in each education group change as the growth of women's work and education shifts the margin of work and college attendance.

occupation may be driven by factors that influence current and future wages[5], I construct a shift-share instrument for gender composition identified from the dramatic increase in relative female labor force participation and educational attainment from 1960 to 2010, measured using the decennial census. Over this time, the female share of educated workers under the age of 35 increased rapidly, particularly for advanced degree holders. This change is unlikely to be attributable to changes in the wages of particular occupations—previous research has attributed this rise in women's work and education to increased access to contraceptives, changing cultural attitudes towards female labor force participation, and changes in the administration of high schools and universities (Goldin, Katz, & Kuziemko, 2006).

This increase in female educational attainment increased the availability of women in all occupations that primarily hired workers with college and advanced degrees. However, because men and women of similar educational attainment tended to work in different occupations, some occupations were more exposed to this change than others. In particular, the rise of female educational attainment and labor force participation induced larger changes in the female fraction of workers in occupations that were equally popular with men and women in 1980, conditional on educational attainment, than in occupations that were disproportionately popular with either men or women in 1980. Relatively equal occupations experienced the largest induced changes because female-dominated occupations had little scope to increase their

[5] Changes in the fraction of workers who are female in an occupation can be related to both endogenous increases and decreases in wage. If occupations that become less demanding see increases in wage, changes in occupations will introduce negative bias to panel regressions. On the other hand, if women's increased education and workforce engagement leads to increases in the fraction female in occupations with increasing skill requirements, changes in occupations will introduce positive bias to panel regressions.

fraction female, while few of the newly-educated female workers sorted into male-dominated occupations.[6]

I use this difference across occupations in exposure to the national trend of increased relative educational attainment of female workers to construct a shift-share instrument for the gender composition of each occupation (Bartik, 1991). This instrument estimates the fraction of available workers in each occupation who would be female if the proportion of workers with each education level, as well as the likelihood that a male or female worker of each education level worked in the occupation, remained fixed to their 1980 levels, while the gender composition of each education level varied over time. Because it is not identified off of changes in the occupational preferences of men and women, this instrument accounts for the endogenous relationship between occupational preferences and wage.

I find strong evidence that increased percent female in an occupation leads to lower average wages for men and women, both in the immediate term and in the longer term. I estimate that a 10 percentage point increase in the female share of an occupation's workforce leads to a 7.7% decline in average male wage, and a 6.1% decline in average female wage, measured contemporaneously. Over the following ten years, the effect grows to an 8.1% decline in average wage for males and an 11.4% decline in average wage for females. Over the following 20 years, the effect on wage falls to a 5.1% decline in average male wage and a 8.8% decline in average female wage. These results are large compared to the cross-sectional relationship between gender composition and

[6] For example, consider an occupation that employs only workers with post-graduate degrees. If men and women with post-graduate degrees were equally likely to work in the occupation, it would have seen an increase in percent female from 15% in 1960 to 59% in 2010, matching the rise in the female share of post-graduate degrees. On the other hand, if post-graduate women were 20 times as likely to work in the occupation as were post-graduate men, the occupation would have been nearly 80% female even in 1960, and nearly 97% female in 2010. Likewise, if post graduate women were 1/20 as likely to work in the occupation as were post-graduate men, the occupation would have been less than 1% female in 1960, but still only 7% female in 2010.

pay—in 2010, when controlling for age and education, a 10 percentage point increase in female share was associated with a 4% decrease in average wage for men and a 5% decrease in average wage for women (Appendix Table A.1).

While this approach faces far fewer threats to identification than do panel approaches that do not instrument for fraction female, there are a few potential sources of endogeneity that must be accounted for in this work. First, because the change in the gender composition of occupations is identified from a shock to female labor supply, I disentangle the effect of labor supply from the effect of gender composition. I do this by constructing a control for the changes in share of the workforce that would be available to an occupation in each year resulting from changes in men's and women's work and education decisions, holding occupation choices constant. This control accounts for the fact that the rise in women's education and work results in higher labor supply in occupations in which a high fraction of workers are highly educated women.[7] Because the occupations with the greatest predicted changes in gender composition are gender-balanced, while the occupations with the greatest predicted change in labor supply are female-dominated, this labor supply control is not collinear with the gender composition instrument.

Second, while the use of this instrument addresses many of the confounding relationships between changes in an occupation's gender composition and changes in wage, shift-share instruments may be confounded by characteristics of occupations that

[7] Increases in the fraction of the workforce consisting of highly educated women would be expected to increase the labor supply of female-dominated occupations so long as demand for educated women did not grow as quickly as supply for educated female workers. This hypothesis is supported by the presence of several supply-side factors that explain the rise in women's work and education, such as delays in childbearing, increased work expectations among women, and greater high-school preparedness (Goldin, Katz, & Kuziemko, 2006). Consistent with a supply-side explanation for the rise in women's work and education, the labor supply index is positively correlated to employment and negatively correlated with wages. If instead changes in demand for female workers drove women's work and education, the labor supply index would be negatively correlated employment and positively correlated with wages.

are correlated with the base-year occupation decisions of men and women. Of particular concern is the possibility that the relative likelihood of female employment in an occupation in 1980 is related to the occupation's skill requirements. In this case, changes in the returns to skills over the last half-century could affect the wages of occupations with large predicted changes in gender composition differently than they affect the wages of occupations with small predicted changes. Work by Autor, Levy and Murnane (2003) and Deming (2017) have identified a few key dimensions of skill that have seen changing returns over this period, with the returns to social skills, math/analytical skills and service increasing and the returns to routine tasks decreasing. While these skills are weakly related to the gender composition of occupations, with social skill requirements highest and math/analytical skills lowest in female-dominated occupations, there is substantial variation in the skill requirements of occupations at each level of relative popularity with female workers. The negative effect of increased percent female on wage is robust to the inclusion of time-varying skill effects, indicating that changing returns to skills identified as important in previous literature do not explain the estimated relationship.

I make several contributions to the literature. First, I contribute to a related literature on the effect on the wages of individuals of working in female-dominated occupations, pioneered by Hirsch and Macpherson (1995) and updated by Addison, Ozturk and Wang (2018).[8] Hirsch and Macpherson measure the effect of changes in occupation gender composition on changes in wages for workers who changed occupations in the CPS. While this analysis finds evidence that workers received lower

[8] Other important work in this literature includes Murphy and Oesch (2015), who examine the wage penalty of changing to a female-dominated occupation in the UK, Germany and Switzerland, and Pitts (2002), who structurally estimates the wage offers of workers in female-dominated occupations, were they to work in male-dominated occupations,

pay in occupations with a higher share of female workers, the difference is mostly explained by measured job characteristics, including training requirements and occupational hazards.

I expand on Hirsch and Macpherson's analysis by examining the effect of gender composition on wage at the occupation level, rather than at the individual level. This differs from Hirsch and Macpherson (1995) in two ways. First, Hirsch and Macpherson analyze workers who change occupation in order to control for the composition of an occupation's workforce. In contrast, I include changes to the composition of an occupation's workforce as part of the causal effect on the occupation's average wage. Second, I examine the wages of a broader group of workers than are considered by Hirsch and Macpherson. Analyses of job switchers only consider workers on the margin between male and female-dominated occupation. This will not capture an effect of gender composition on the wage of an occupation, because a decline in the wages of an occupation would be expected to shift marginal workers out of that occupation while lowering the wages of non-marginal workers. As a result, female representation can causally lower wages in an occupation without generating a premium for individual workers for leaving the occupation.

Next, I contribute to the growing literature examining the effect of changes in gender composition on the characteristics of occupations. Previous research has hypothesized that the entry of women into occupations causes declines in an occupation's prestige (Goldin C. , 2014), a hypothesis supported by the exit of men from occupations that pass a threshold of percent female (Pan, 2010). In addition, Goldin and Katz (2011, 2012) find that the gender composition of an occupation coevolves with returns to long hours and rates of independent ownership, and they speculate that some of that coevolution may be due to the effect of gender composition on occupation

characteristics. The large, negative effect of gender composition on wage found in this paper suggests that these hypothesized effects of gender composition on occupations may be large and have significant wage consequences.

The rest of this paper is organized as follows: Section 1.2 introduces theory suggesting that gender composition of an occupation can affect its wage. Section 1.3 describes the data used in this analysis. Section 1.4 describes the empirical strategy of the paper, Section 1.5 describes the results, Section 1.6 describes tests of the mechanisms of effect, and Section 1.7 concludes.

Section 1.2: Theory

Gender composition can affect the average wage of an occupation through three primary channels: changes in the distribution of characteristics of workers in an occupation, changes in the characteristics of the occupation, and changes in the way that worker characteristics are compensated. That is, a change in the percent of workers who are female can cause workers to enter or exit an occupation (e.g. highly skilled workers might exit the occupation), can cause an occupation to change amenities (e.g., the occupation might require shorter working hours), or can change the returns to worker characteristics (e.g. returns to particular skills could increase or the cost of career interruptions could decrease). In this analysis, all three of these channels contribute to the causal effect of interest.

Those causal channels may in turn be activated by several underlying economic mechanisms. I will divide these mechanisms into two broad categories: effects on wage through prestige/perceptions of the occupation, and effects on wage through amenities/characteristics of the occupation.

Mechanism 1: Prestige/Perceptions:

Declines in wage as a result of increased fraction female may be the result of negative perceptions of women's abilities or work engagement influencing the perception of women's occupations. For example, Goldin (2014) shows that if observers outside an occupation believe that female workers have lower abilities on average than male workers, and they cannot perceive the true skill requirements of an occupation, male dominated occupations will be perceived to require more skill than female-dominated occupations. An increase in female representation in an occupation may therefore result in a decline in perceived difficulty (prestige) of the occupation among individuals not working in the occupation.

This can affect wages of workers both through direct channels, like declines in demand for labor in the occupation, or through changes in the composition of the workforce. Workforce composition may change if, in addition to its psychic benefits, prestige influences the ability of a worker to get a higher-paying job in the future. Because workers change occupation frequently, with between 20-30% of respondents in the NLSY reporting a change in occupation and firm each year (Sullivan, 2010), the signal of ability sent by current occupation may be an important factor in determining future wages. If lost opportunities for career advancement are costlier to higher ability workers, this will result in selection of lower-ability male and female workers into occupations that see increasing shares of female workers. As a result, even if compensating differentials increase wages in the occupation conditional on ability, average wage may fall due to a shift in the composition of the workforce toward lower-ability or lower-ambition workers.

The effect of a decline in prestige on the wages in an occupation may be long-lasting. Basu (2017) notes that if many agents in society benefit from giving preferential treatment to groups that are preferred by others, discrimination can persist in

equilibrium. For example, suppose that technology companies can be run either by engineers or by marketers. An increase in the percent of marketers who are female leads investors and potential employees to conclude that marketing is a less demanding and less prestigious occupation than is engineering. If, as a result, employees prefer to work for companies run by engineers and investors prefer to invest in companies run by engineers, engineers will be favored for leadership positions over marketers. Once this equilibrium is established, employees will find it optimal to preferentially work for engineers, because engineers are better at securing funding than are marketers. Investors will also find it optimal to invest in engineer-run companies, because engineers can secure better employees. Equilibrium outcomes of this sort could mean that occupations that see an increase in the share of female workers become unable to attract ambitious workers in the long run.

Mechanism 2: Changing amenity value of the occupation:

Increased female representation in an occupation can also change wages by changing the optimal provision of job amenities to workers. The provision of job amenities may change with changes in gender composition because women have different average preferences towards workplace amenities than men. In particular, women on average pay higher costs for long and inflexible working hours than do men (Goldin C. , 2014), (Mas & Pallais, 2017), are less willing to apply for jobs with competitive compensation schemes (Flory, Leibbrandt, & List, 2015), and are less willing to accept jobs that risk fatality (DeLeire & Levy, 2004).

In many circumstances, work can be organized in a way that allows shorter and more flexible hours, reduced competition and/or less accident risk at the cost of lower productivity or higher costs. For instance, Goldin and Katz (2011) find that small veterinary practices can reduce the costs of flexibility in hours by referring emergencies

to regional veterinary hospitals, rather than by keeping staff "on call" outside of normal working hours. In this case, a veterinary practice will choose to refer emergencies if the cost of lost productivity is lower than the benefit of lower wages, due to the amenity value of flexible hours. Because female veterinarians are less likely than male veterinarians to work overtime hours (Goldin & Katz, 2011), referring cases becomes more profitable as the percent of veterinarians who are female increases. In many cases, such reorganizations reflect the gender composition of occupations as much as they do particular firms—as the fraction of veterinarians who are female increases, regional hospitals will receive more referrals, become more prevalent, and thus suffer fewer disadvantages relative to on-call hours. Indeed, as women began to dominate the veterinary profession—making up 80% of recent veterinary school graduates—emergency services have shifted to regional hospitals (Goldin & Katz, 2011).

This process can occur for any amenity that is valued more by female than by male workers on average, and would be expected to lower average wages. Increased amenities will lower wages conditional on worker characteristics by increasing costs and/or lowering productivity. In addition, it will make the occupation more appealing to workers with a high value of the amenity (i.e. those who face high costs from long hours, competition etc.) and less appealing to workers with a low value of the amenity, due to decreased wages. If workers who face greater costs from competitive pay structures, long hours and other amenities are on average less productive or less ambitious than those who face low costs, lower productivity workers will sort into an occupation as the workplace amenities improve.

It is important to note that these two mechanisms are not mutually exclusive, and in fact may be mutually reinforcing. A decline in an occupation's prestige may result in a change in the composition of a workforce toward less ambitious workers, who

in turn may demand flexible working hours, less risky compensation schemes and similar amenities. Likewise, higher provision of amenities in an occupation may result in lowered perceptions of the occupation's difficulty, and thus diminished prestige.

Section 1.3: Data

The main analysis in this paper is conducted using the IPUMS microdata files of the Decennial Census (1960-2000) and the American Community Survey (2009-2014) (Ruggles, Genadek, Goeken, Grover, & Sobek, 2015). Supplementary analysis is performed using the Panel Study of Income Dynamics (1968-2016) and the Current Population survey (1968-2016). These data are described in detail in the Data Appendix, but I focus on a few important issues here.

First, I choose the years 1960 to 2010 in order to maximize sample size while maintaining consistent ten-year intervals between included years. I exclude the 1950 census because the sampling technique used to measure occupation was non-random within households selected for the long-form census. In particular, the 1950 census asked about occupation only for one member of each household, selected by the household. While it is possible to reweight this sample in order to make it nationally representative, doing so cannot account for correlations between the likelihood that the selected respondent is female and the occupation of the female household member.

Second, I define occupation using a modification of the 1990 occupation codes constructed by IPUMS. These codes are constructed using crosswalks to earlier and later occupation coding schemes, as described by Myer and Osborne (2005). I deviate from the 1990 harmonized codes in the 2000 and 2010 census by using the gender-specific occupation classifications provided by Blau, Brummund and Liu (2012). Blau et. al. observed that because the most prevalent 1990 codes for each 2000 code differ

for male and female workers, gender segregation of occupations appears to decrease in 2000 when measured with harmonized 1990 occupation codes. To address this, Blau et. al. define a gender-specific crosswalk between the 1990 and 2000 codes by choosing the most prevalent 1990 occupation separately for women and men for each 2000 occupation code.

Finally, I include all occupations in my analysis, regardless of size, and do not weight occupations by size. There are two reasons that one might weight occupations in this context. First, as discussed in Solon, Haider and Woolridge (2013), the measurement error associated with mean wage in each occupation is decreasing in the size of the occupation, causing heteroskedasticity in the error term. Were errors uncorrelated within group, weighting occupations by the square root of their sample size would address this heteroskedasticity. However, Solon, Haider and Woolridge (2013) note that when errors are clustered within occupation, as they are in this analysis, unweighted estimates may be more precise. I performed a simulated power analyses, considering four ways of addressing unequal size of occupations: no weights or exclusions, weighting by sample size, weighting by the square root of sample size, and eliminating occupations represented by fewer than 100 workers. Of these, analyses with no weights or exclusions had the most power. Secondly, weighting estimates by the number of workers in each occupation would make the estimated effect representative of the effect in the average worker's occupation, whereas unweighted estimates represent the effect in the average occupation. Because occupations, rather than workers, are the subject of this analysis, I perform an analysis that is representative of the average occupation. As a robustness check, I present estimates weighted by the number of workers in each occupation Appendix Table A.13. when

weighting by number of workers, the estimated effect of fraction female on wage is smaller than when not weighting, but still large and statistically significant.

Section 1.4: Empirical Strategy

In order to motivate the approach taken in this paper, I consider a hypothetical experiment testing the effect of gender composition on wage. One such experiment might divide occupations into treatment and control, but would retire some random subset of the male workers in the treated occupations and replace them by cloning a random subset of female workers. The experimenter would then measure average wages for male and female workers in this occupation, including in the average workers who enter the occupation after the cloning experiment and excluding in the average workers who exit the occupation after the cloning experiment.

This experiment has a few important properties. First, it would effectively replace male workers whose characteristics are distributed typically for males working in the occupation with female workers whose characteristics are distributed typically for females working in the occupation. As a result, an effect on wage in this experiment could come through direct effects of gender, such as declines in prestige for occupations with a large female workforce, but could also come through indirect effects of gender-linked characteristics such as demand for amenities. Second, because the experimental effects include entry to and exit from the occupation, the estimated coefficient is an effect on the average wage of the occupation, not on the expected wage of a worker with a given set of characteristics.

It is important to note that this experiment measures only one of several parameters that could be considered an effect of the gender composition of an occupation on wage. Because this experiment changes gender composition without

changing the preferences for or barriers to an occupation, it may have quite different effects than would an experiment that provided a quota for female workers in an occupation, that lowered discrimination against women in an occupation, or that implemented affirmative action in education that is a prerequisite for work in the occupation. In particular, these alternative experiments are likely to include labor supply effects through the overcrowding channel (Bergmann, 1974) that are not included in the effects measured in the primary experiment. Yet another alternative might be to transform the sex of some subset of workers in treated occupations from male to female, without changing the preferences, beliefs or constraints of the workers at all. Doing this would present an effect of gender composition absent the effect of any differences between male and female workers within an occupation.

1.4.1: OLS Panel Regression:

My goal is to replicate the primary experiment described above by examining the effect of changes in the fraction of workers in an occupation who are female in one census year on wages in future census years. Because the margin of work and college attendance is changing for men and women over this period, the unobserved abilities of workers of each education level may have changed (Lovenheim & Reynolds, 2011). I address this by estimating the effect on wages of workers aged 45-65, who I label "Incumbent" workers, from changes in the gender composition of the occupation driven by the education and work decisions of workers aged 22-35, who I label "Newcomer" workers. I estimate the following regressions, for male and female workers, indexed by gender g, and for zero, ten and twenty-year time-lags, indexed by k:

$$W_{g,j,t+k} = \alpha + \beta_{1k} f_{j,t} + \beta_{2kf} W_{f,j,t} + \beta_{2km} W_{m,j,t} + \delta_{fk} X_{f,j,t} + \delta_{mk} X_{m,j,t} + \gamma_{jk} + \sigma_{tk} + \varepsilon_{g,j,t+k} \quad (1)$$

Where $W_{g,j,t}$ is the log mean wage of workers aged 45-65 with gender g in occupation j at year t, $f_{j,t}$ is the proportion female of workers aged 22-65 in occupation j, and $X_{m,j,t}$, $X_{f,j,t}$ are the mean workforce characteristics male and female incumbent workers respectively. In this regression, β_{1k} measures the effect of fraction female in year t on log mean wage in year $t+k$, and includes the effects of changing workforce composition (other than characteristics in $X_{f,j,t}$), changing returns to worker characteristics and changing occupation characteristics. I highlight a few key characteristics of this regression below.

First, because β_{1k}includes the effect of fraction female on workforce composition and occupation characteristics, controls for time-varying occupation characteristics and worker characteristics may attenuate estimates by capturing changes in those characteristics induced by the change in gender composition. As a result, I include in the controls $X_{j,t}$ only a limited set controls for the age and the distribution of education for workers in the occupation, because these are the occupation characteristics most likely to be confounded with the instrument. In particular, I control for the percent of workers between the ages of 45-55, the percent of male and female workers aged 55-65, and the average age of the occupation among workers aged 45-65. I also control for the percent of male and female workers who have less than a high school education, a high school education, an associate's degree, a bachelor's degree, and an advanced degree.

Next, I estimate a linear relationship between fraction female and log wage. I do this for a mix of practical and theoretical reasons. A linear relationship is supported by theory because several potential mechanisms suggest that wages should be influenced by changes to the percent of female workers, rather than to the growth rate of female employment. In particular, the returns to amenities that operate as public goods in an occupation depend on the sum of utilities across all workers, so differences in the value

of those goods to men and women should be weighted by their raw percentages. The proper functional form of the relationship between percent female, prestige and wage is less clear, but absolute changes in percent female are observable to those outside an occupation in a way that proportional changes may not be.

As a practical matter, examining changes in the percent female avoids measurement error. The decennial census is a very large dataset, sampling 1% of the US population in year 1960, 1970, 2000 and 2010, and sampling 5% of the US population in 1980 and 1990. However, construction of the shift-share instrument (described in detail in Section 1.4.2) is data intensive, and places a great deal of weight on single individuals, particularly in heavily male-dominated and heavily female-dominated occupations. This is particularly true for occupations that were small in the base year. An occupation with 200 members that is 2% female would be represented in the base-year by only four female workers. As a result, measurement of the educational composition of female workers in that occupation is necessarily extremely noisy. While the predicted percent changes in such an occupation would be small regardless of noise in the number of females and educational distribution of those females, changes in log percent female are large and highly variable for such an occupation.

Finally, I examine changes in wage over the course of 20 years because occupations are likely to evolve slowly in response to changes in gender composition. Firms may take time to provide amenities to their workers after demand for those amenities increases. Furthermore, the wage consequences of those changes, due both to selection into and out of the occupation and to the productivity losses associated with occupational amenities, may differ in the immediate-term from the long-term. Likewise, negative prestige consequences of increased female representation in an occupation may take time to arise, as those outside of the occupation observe the change in the

workforce, and may change in importance over time as observers get additional information about the occupation.

1.4.1.1 Potential Confounders

Because there are several important sources of endogeneity between fraction female and wage, equation 1 does not identify the causal effect of fraction female on wage. A few particularly important sources of endogeneity are described below.

Female Labor-Force Engagement: Because female attachment to the labor force grew considerably from 1960-2010, and because women dramatically increased their education levels relative to men, occupations that attract new and highly educated workers would be expected to see increases in the representation of female workers, relative to occupations that are less attractive to new and highly-educated workers. Because new, educated workers are likely to choose occupations that are expected to experience high growth and a positive wage outlook, this will produce a positive relationship between changes in percent female and average wage. In addition to the mechanical effect—more new, educated workers leading to more female workers because new, educated workers are disproportionately female—we might expect that discrimination against female workers is less costly in occupations that are not experiencing growing labor demand and tight labor markets. In addition, women's high elasticity of labor supply relative to men (McClelland, Mok, & Pierce, 2014) implies that changes in the wage level of an occupation will have a larger effect on the labor force participation of women trained to work in the occupation than on the labor force participation of men trained to work in the occupation, causing positive correlation between wage and percent female.

Returns to Skills Within an Occupation: Because men and women have different average skills, changes in the returns to skills may also affect both gender composition and wage growth. If the returns to skills that are more prevalent among male workers decrease in an occupation (for instance, if physical strength becomes less important), wages will fall and female representation will rise. Meanwhile, if the returns to skills that are more prevalent among female workers increase in an occupation (for instance, if social skills become more important (Deming, 2017)), wages and female representation will rise. This bias could be both positive or negative, depending on the relative importance of and changes in disproportionately male and disproportionately female skills.

Aggregation of Sub-Occupations: Finally, because occupation definitions necessarily aggregate over several types of work, changes in the relative size of male-dominated and female-dominated sub-occupations can change both gender composition and wages. For example, secretaries at manufacturing firms are more likely to be male and are more highly paid than are secretaries at other firms. A decline in the percent of secretaries who are manufacturing secretaries will thus increase the percent female among secretaries and decrease wages, even if it is not associated with any change in wage or gender composition among manufacturing secretaries or among other secretaries.

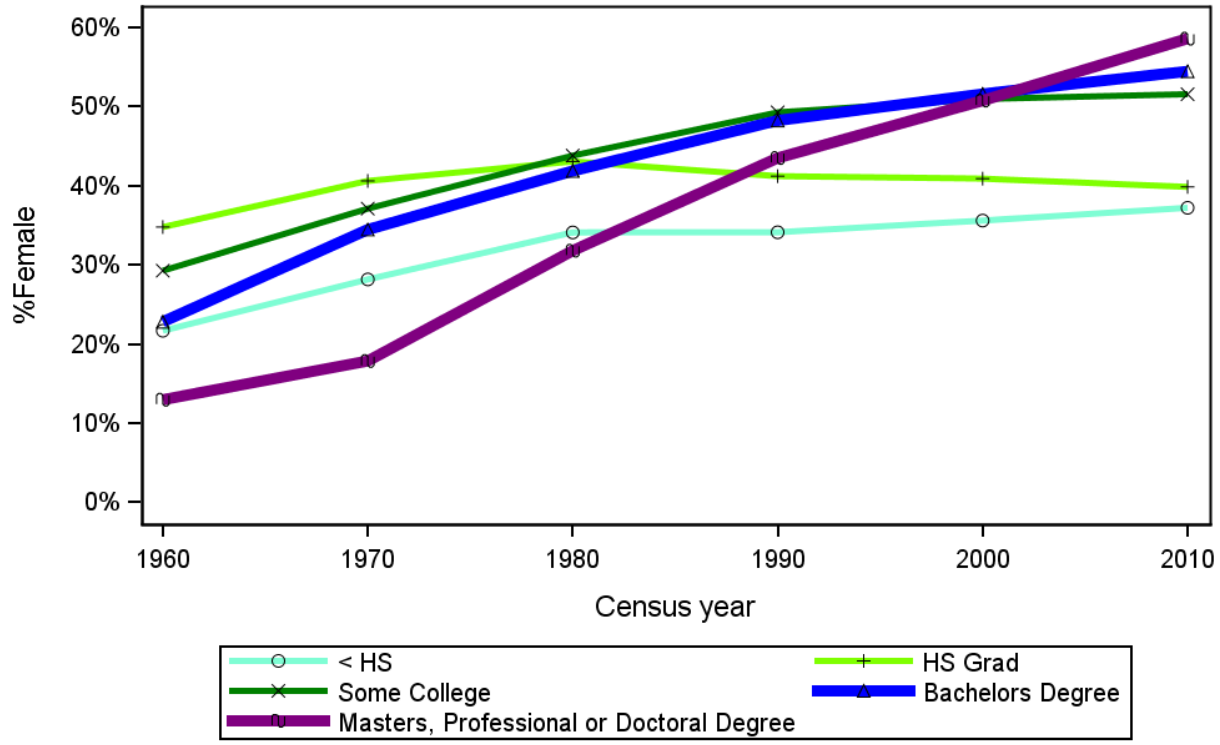
1.4.2: Instrumenting for Gender Composition

1.4.2.1: Definition of Primary Instrument

I address these cofounders by exploiting the increase in the female fraction of the workforce brought about by the rapid increase in women's relative educational attainment and labor force participation from 1960 to 2010. As shown in Figure 1.1,

the percent of college degree and advanced degree holders who are female increased dramatically among workers aged 22-35, with just over 11% of advanced degree holders in the workforce female in 1960, compared to nearly 60% in 2010.

Figure 1.1: Gender Composition of Full-time workers aged 18-35 by highest degree



Note: This figure shows the percent of US workers of each education type aged 22-35 who are female in each census year from 1960-2010. Definitions of Education Type are available in the data appendix.

Were all occupations to remain fixed at some point in time in their relative attractiveness to male and female workers with each education level, this change in the educational attainment of female workers relative to male workers would increase the availability of female workers in all occupations, particularly in occupations that hire workers with college degrees or advanced degrees. However, the extent of the implied change would depend on the relative prevalence of the occupation among males and

females of each education level. Occupations like physics or nursing, that are predominantly chosen by workers of one gender, would have small induced changes in gender composition, while occupations like psychology, which are chosen by both male and female workers, would have larger induced changes. The occupations for which these changes are largest and smallest over the full sample period are shown in Appendix Tables A.1 and A.2.

The occupations that experienced the largest change in gender composition induced by the instrument are diverse. Social sciences are highly represented, with Psychologists, Economists and Social Scientists (not otherwise classified) all among the ten occupations with the highest predicted change in fraction female. Relatively male-dominated segments of the education field are also well represented, with High School Subject Instructors and Managers in Education in the top 10. Therapy and Social Work also experienced high predicted growth in fraction female, as did writers and authors, editors and reporters and artists/entertainers. With the exception of artists/entertainers, workers in these occupations were highly educated in 1980, with 70-95% of workers holding at least a college degree. They are all people-oriented, involving either close contact with other people or careful study of human behavior. However, they vary substantially in other skills and attributes, including occupations with different firm structures, risk and job requirements.

Occupations that experienced the smallest change in gender composition induced by the instrument are dominated by engineering occupations, but also include particularly male-dominated and particularly female-dominated healthcare occupations. Six of the ten occupations with the smallest predicted change in fraction female are engineering occupations, as are seven of the top 20. All of these engineering occupations had few female workers in 1980 and required high levels of education.

Other occupations with a low predicted change in fraction female are quite diverse, however, including registered nurses, dieticians, physicists and dentists. While registered nurses and dieticians were overwhelmingly female in 1980, dentists were overwhelmingly male—only 5% of dentists were female in 1980. The low predicted changes in dentistry and optometry contrast with other health diagnosing occupations—pharmacists, physicians and veterinarians were all more prevalent occupations for women than were dentistry and optometry in 1980, and were all predicted to have average or above-average increases in fraction female.

I define a shift-share instrument (Bartik, 1991) to capture these induced gender compositions by fixing in 1980 the fraction of workers in each occupation who have each of five education types, defined by highest degree attained: less than high school, high school, some college, bachelor's degree, and master's degree or higher. In addition, I fix the fraction of men with each education type working in the occupation and the fraction of women with each education type working in the occupation. The instrument gives the fraction of workers in occupation j who would be female were these base-year fractions fixed, but the gender composition of each education level varied from year to year. I define the instrument as:

$$\widetilde{f_{J,t}} = \sum_A \gamma^j_{a,1980} * \left(\frac{\omega^F_{aj,1980} * f^{\bar{J}}_{a,t}}{\omega^F_{aj,1980} * f^{\bar{J}}_{a,t} + \omega^M_{aj,1980} * \left(1 - f^{\bar{J}}_{a,t}\right)} \right) \quad (2)$$

Where 1980 is the base year, $\omega^F_{aj,1980}$ is the fraction of females with educational attainment a in occupation j in 1980, $\omega^M_{aj,1980}$ is the fraction of males with educational attainment a in occupation j in 1980, $f^{\bar{J}}_{a,t}$ is the female fraction of workers with educational attainment a in year t, excluding those employed in the same occupation as occupation j, and $\gamma^j_{a,1980}$ is the fraction of workers in occupation j with educational

attainment a. This can be simplified by defining $\Omega_{a,j,1980} = \frac{\omega^{M}_{aj,1980}}{\omega^{F}_{aj,1980}}$, and $F^{\bar{J}}_{a,t} = \frac{f^{\bar{J}}_{a,t}}{1-f^{\bar{J}}_{a,t}}$. I can then define:

$$\widetilde{f_{J,t}} = \sum_{A} \gamma^{j}_{a,1980} * \left(\frac{\Omega_{a,j,1980} * F^{\bar{J}}_{a,t}}{\Omega_{a,j,1980} * F^{\bar{J}}_{a,t} + 1} \right) \tag{3}$$

To illustrate the predictions of this instrument for male-dominated, female-dominated and balanced occupations, consider the examples of physics, psychology and nursing. Table 1.1 gives the fixed base-year values for each of these occupations. Physics was a far more common occupation for males of each education level than for females of the same education level in 1980, with a man holding an advanced degree being more than 10 times as likely to be a physicist than was a woman holding an advanced degree in 1980. In contrast, registered nursing was a far more common occupation for females of each education level than for males, with a woman holding an associate's degree more than 30 times as likely to be a registered nurse than was a man holding an associate's degree. In contrast, psychology was similarly common among men and women of each education level, with women a woman holding an advanced degree 1.7 times as likely to be a psychologist as was a man holding an advanced degree.

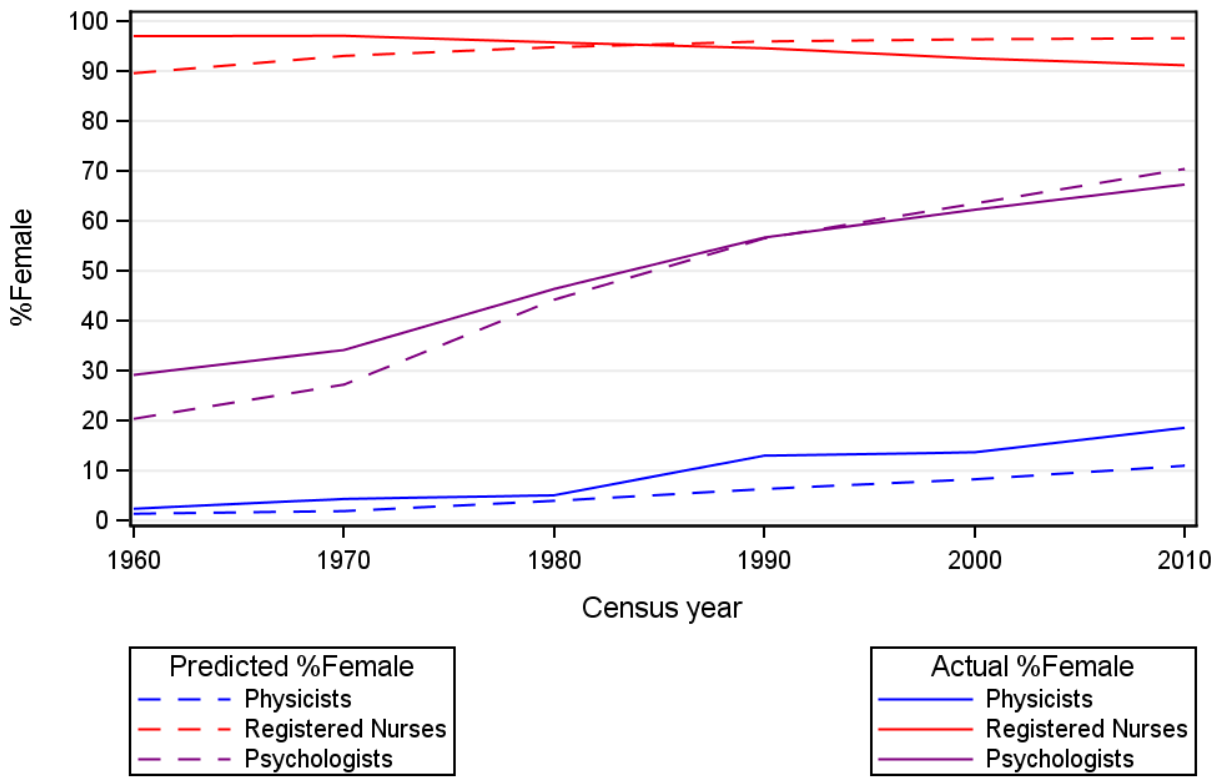
Table 1.1: Base-Year Fixed Components for Three Example Occupations:

	% of Workers With Ed	Workers / 10,000 Females	Workers /10,000 Males	Female / Male Ratio
Physicists and astronomers				
1: No Degree	0.53	0.03	0.11	0.31
2: HS Degree	3.18	0.07	0.42	0.16
3: Associates Degree	7.56	0.30	1.55	0.19
4: College Degree	17.90	0.57	5.03	0.11
5: Advanced Degree	70.82	3.18	36.64	0.09
Psychologists				
1: No Degree	1.22	1.02	0.39	2.62
2: HS Degree	2.55	1.05	0.51	2.07
3: Associates Degree	4.55	2.68	2.45	1.10
4: College Degree	15.21	14.66	8.87	1.65
5: Advanced Degree	76.47	177.96	104.55	1.70
Registered nurses				
1: No Degree	1.27	16.39	1.13	14.53
2: HS Degree	7.36	48.57	2.43	20.02
3: Associates Degree	55.09	739.75	24.34	30.39
4: College Degree	30.03	568.05	26.69	21.28
5: Advanced Degree	6.26	345.46	24.35	14.19

*Notes: % of workers with ed gives the percent of workers in occupation j with degree a in 1980, workers/10,000 Females gives the percent of women with degree a working in occupation j, and workers/10,000 males gives the percent of men with degree a working in occupation j. Female/Male Ratio gives (Workers/10,000 Females)/(Workers/10,000 Males)

These differences in base-year characteristics mean that these three occupations vary substantially in their exposure to the changing gender composition of education groups shown in Figure 1.1. These differences are illustrated in Figure 1.2, which shows the percent female in each occupation predicted by the instrument and the actual percent female for these three occupations in each year from 1960 to 2010. As the fraction of degree holders who are female increases, the predicted percent female for each of these occupations increases, but the increase is much larger for psychology than for nursing or physics.

Figure 1.2: Predicted and Actual percent female for Three Example Occupations:



Note: This figure shows the percent female among workers in three occupations, as well as the instrumented percent female for young workers in those occupations

The differences between the predicted changes in physics, nursing and psychology illustrate the identification of this paper. While these occupations differ from each other in many ways, the large predicted increase in the fraction of psychologists who are female, relative to the smaller predicted increases in the fraction of nurses and psychologists who are female, is not due to changes in psychology, in nursing or in physics. Instead, it is a consequence of the interaction of psychology's preexisting gender composition with broad changes in the role of women. As a result, these changes are far less likely to be confounded with changes in demand for these occupations, in

the difficulty of work in these occupations, or in the returns to skills in these occupations than are actual changes in gender composition.

1.4.2.2: Additional Controls

The construction of this instrument produces two sources of endogeneity that require the introduction of controls. First, I control for base-year gender composition of each occupation interacted with year. Specifically, I include controls for $\gamma^{j}_{a,1980} * (\acute{t} = t)$, for each level of educational attainment a and year t, where $\gamma^{j}_{a,1980}$ is the fraction of workers in occupation j with education attainment a in 1980 and $\acute{t}$ is a particular census year. This controls for the time-varying effect of the base-year education composition on wage. This control is necessary because occupations that hire a larger fraction of workers from high education levels in the base year have larger implied increases in percent female. Because the returns to education change over time, this leads to a relationship between $\widetilde{f_{J,t}}$ and $W_{g,j,t}$ that is not related to the gender composition of the occupation.

Second, I control for the changes in labor supply across occupations brought about by the rise in women's relative education and labor force participation. Because this rise increased the fraction of the workforce available to work in occupations that hire a large fraction of highly educated women, it might be expected to depress wages in occupations that draw more female than male workers. This effect of rising women's education and labor force participation is theoretically distinct from the effect of gender composition because it depends on the number of workers available in the occupation, not whether those workers are male or female.

I control for the effect of education and work decisions on labor supply by constructing and controlling for a labor supply index. This index estimates the fraction of the workforce that would work in occupation j in year t were men and women's education and labor-force participation decisions to change but were their occupation decisions, conditional on education and labor-force participation, to remain constant. This labor supply index is defined below:

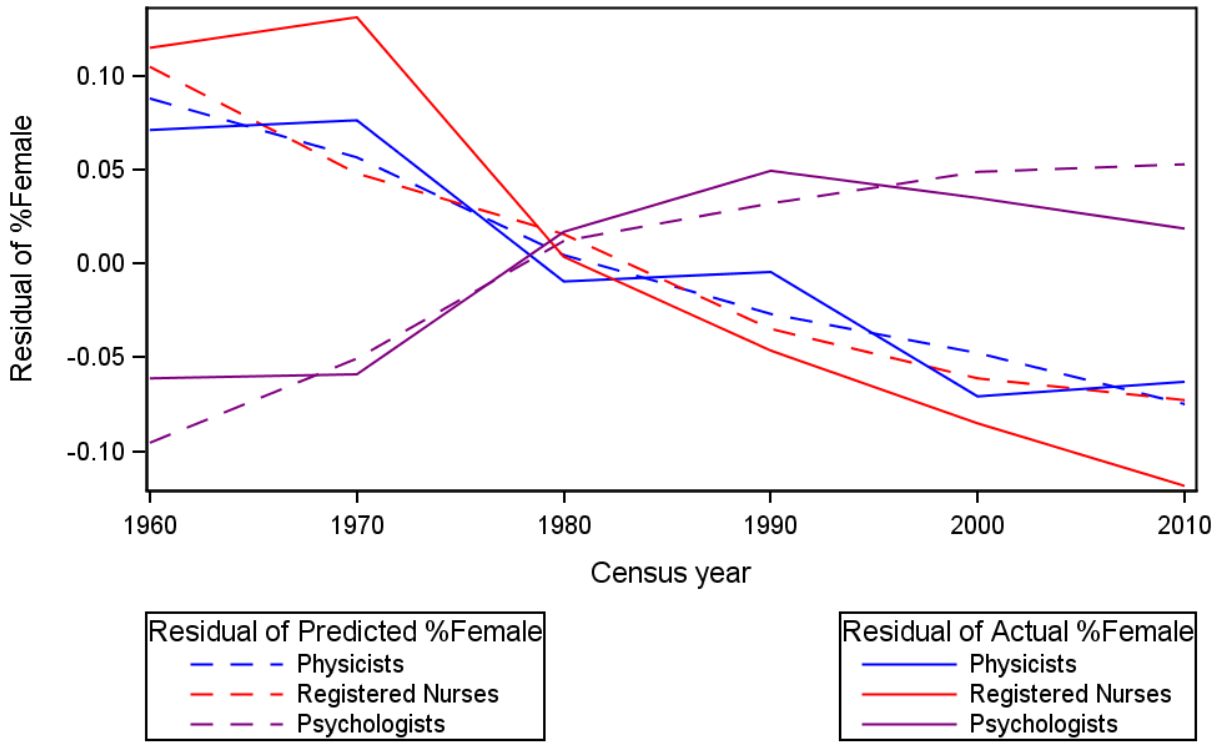
$$\widetilde{l_{j,t}} = log\left(\sum_A \gamma^j_{a,1980} * \left(\omega^F_{aj,1980} * k^{\bar{j}}_{f,a,t} + \omega^M_{aj,1980} * k^{\bar{j}}_{m,a,t}\right)\right) \qquad (4)$$

Where $k^{\bar{j}}_{f,a,t}$ is the fraction of the labor force at time t composed of females with educational attainment a, and $k^{\bar{j}}_{m,a,t}$ is the fraction of the labor force at time t composed of males with educational attainment a. Like the gender composition index $\widetilde{f_{j,t}}$, this index holds the education composition of the occupation and the occupation choices of men and women in the workforce with each education level a fixed. However, rather than taking the ratio of the implied female workforce to the implied total workforce, it adds the implied female workforce to the implied male workforce. I take the log of this index in order to interpret coefficients on labor supply as elasticities. The log of the labor supply index is also more strongly correlated with wage than is a linear index normalized to 1980 labor supply, but the choice of functional form has small effects on the estimated effect of fraction female on wage.

The occupations with the largest induced changes in labor supply over the sample period are not the occupations with the largest induced change in fraction female. To show this, I depict the relationship between base-year gender composition by educational attainment and the instrument graphically in Figure 1.3. This figure defines a summary statistic, "Weighted Average Female/Male Popularity by Education", which represents the average ratio of $\omega^F_{aj,1980}/\omega^M_{aj,1980}$, weighted by $\gamma^j_{a,1980}$. I compare

changes in the instrument and to labor supply over the whole sample period in male-dominated occupations (occupations with a low weighted log gender ratio), in gender-balanced occupations (occupations with a weighted log gender ratio near 0) and in female-dominated occupations (occupations with a high weighted log gender ratio). This figure shows that occupations with the highest predicted changes in the instrument have balanced gender ratios, meaning that they are similarly popular among men and women of each education level. In contrast, changes in induced labor supply are highest in occupations with high ratios, indicating that they are disproportionately popular among female workers.

Figure 1.3: Residual of Predicted and Actual percent female for Three Example Occupations



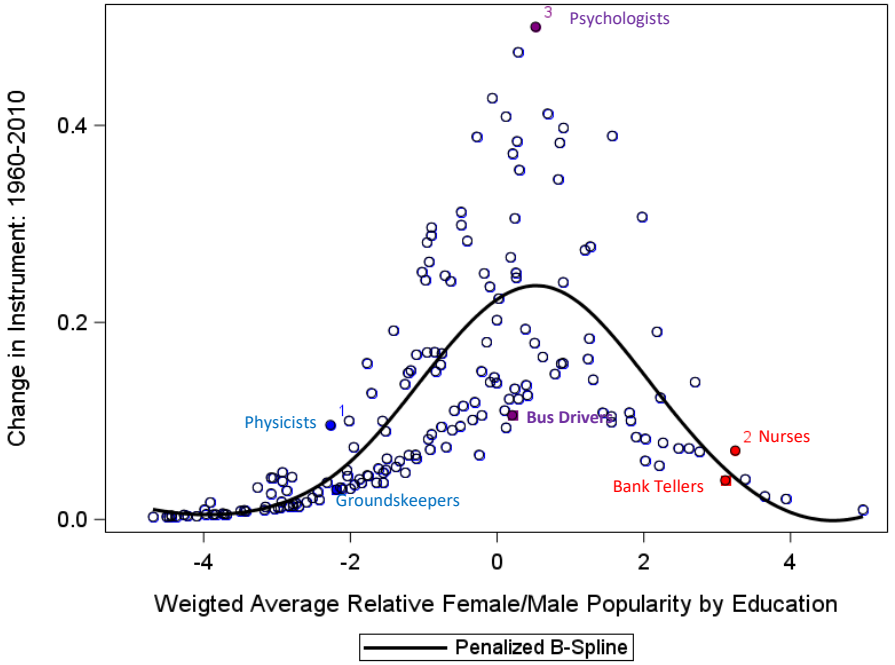
Note: This figure shows the residual of percent female and instrumented percent female for workers in three occupations after controlling for occupation and year fixed effects and for controls described in the text.

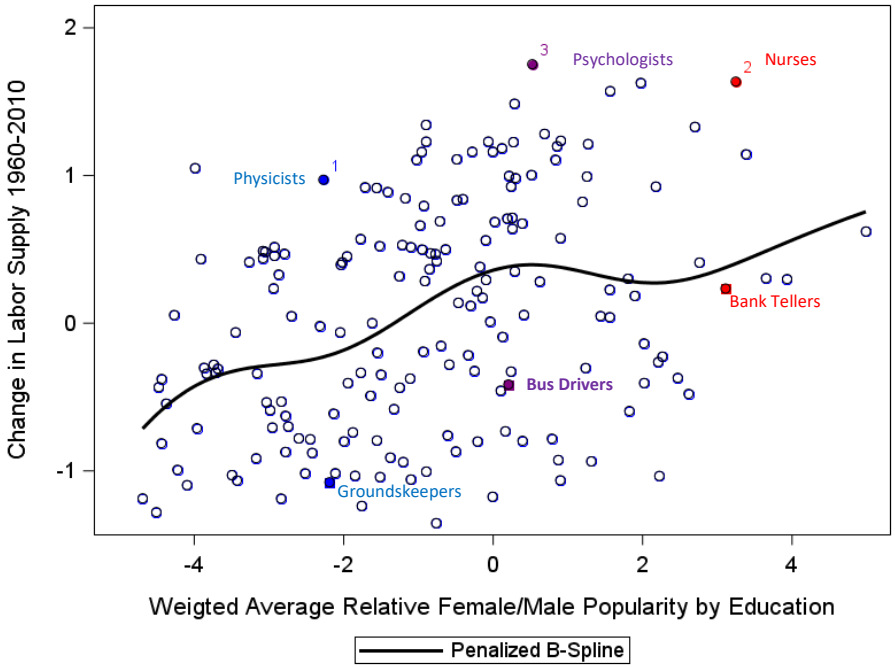
This difference in distribution is crucial to disentangling the effect of fraction female from the effect of labor supply. A wage effect operating through labor supply should be greatest for female-dominated occupations, moderate for gender-balanced occupation, and lowest for male-dominated occupations. In contrast, a wage effect operating through fraction female should be greatest for gender-balanced occupations and low for both male-dominated and female-dominated occupations.

After including these controls, the differences in predicted changes in gender composition for physicists, nurses and psychologists remain significant. Figure 1.4 shows the residuals of the instrumented and actual gender composition of the three

example occupations when including educational attainment and labor supply controls, along with occupation and panel fixed effects and time-varying age and education controls. While nursing and physics are predicted to see declining percent female relative to the trend in gender composition of similarly educated occupations, psychology is predicted to see increasing percent female relative to similarly educated occupations. These predictions match the actual gender composition of these occupations.

Figure 1.4: Relationship Between Instrument, Labor Supply and Relative Female/Male Popularity by Occupation





Notes: Weighted Average Relative Female/Male Popularity Represents the ratio of the percent of female workers of each education level in an occupation to the percent of male workers of each education level in an occupation, weighted by the percent of workers of the education level in the occupation. These graphs show the relationship between relative popularity and change from 1960-2010 in the instrument, induced total labor supply, induced male labor supply and induced female labor supply.

1.4.2.3: Regressions of Interest:

I generate my main results by running a two-stage least squares analysis of equation 1, using the instrument defined in equation 4. The regression equations are below, for males and females, indexed by k, and for 0, 10 and 20 year time-lags, indexed by k.:

First-Stage:

$$f_{j,t} = \beta_{1k}\widetilde{f_{j,t}} + \beta_{2kf}W_{f,j,t} + \beta_{2km}W_{m,j,t} + \beta_{3k}\widetilde{l_{j,t}} + \delta_{fk}X_{f,j,t} + \delta_{mk}X_{m,j,t} + \gamma_{jk} + \sigma_{tk} + \varepsilon_{j,t} \quad (5)$$

Reduced-Form, Simultaneous:

$$W_{g,j,t+k} = \vartheta_{1k}\widetilde{f_{j,t}} + \vartheta_{2kf}W_{g,j,t} + \vartheta_{2km}W_{m,j,t} + \vartheta_{3k}\widetilde{l_{j,t}} + \vartheta_{4kf}X_{f,j,t} + \vartheta_{4km}X_{m,j,t} + \upsilon_{jk} + \varsigma_{tk} + \pi_{g,j,t} \quad (6)$$

Structural Equation:

$$W_{g,j,t+k} = \theta_{1k}\widehat{f_{j,t}} + \theta_{2kf}W_{g,j,t} + \theta_{2km}W_{m,j,t} + \theta_{3k}\widetilde{l_{j,t}} + \theta_{4kf}X_{f,j,t} + \theta_{4km}X_{m,j,t} + \eta_{jk} + \lambda_{tk} + \mu_{g,j,t} \quad (7)$$

1.4.2.4: Exogeneity of the Instrument:

As with all shift-share instruments, the exclusion restriction for this instrument depends on the panel structure of data and the presence of occupation and year fixed effects. The exclusion restriction for the instrument is that the instrument is uncorrelated with the error term $\mu_{t,j}$ of structural equation (7). This exclusion restriction indicates that there is no correlation between the instrument and present or future wage after controlling for year fixed-effects, occupation fixed-effects, education distribution at time t, age distribution at time t, induced labor supply, and education distribution in 1980 interacted with year, other than through $f_{j,t}$.

Goldsmith-Pinkham Sorkin and Swift (2018) (GPSS) note that this exclusion restriction holds in two cases. In the first case, the base-year fixed shares must be exogenous to changes in wage over time (equivalent to a common-trends assumption in

a difference-in-difference estimation). In the second case, the time trends must be exogenous to dispersion in wages across occupation. GPSS formalize this argument for a traditional Bartik instrument, where the instrument is defined as the sum of industry shares multiplied by industry growth rates. This instrument is algebraically equivalent to instrumental variables estimated using the General Method of Moments (IV GMM), where the moment conditions are that the base-year shares multiplied by year indicators must be uncorrelated with the structural error term. The instrument used in this paper interacts base year components (here the ratio of male to female employment in occupation j given educational attainment a) nonlinearly with time trends (here the ratio of female to male workers with educational attainment a). As a result, this instrument is not mathematically equivalent to IV GMM.

However, the exogeneity of the instrument still relies on the assumption that base-year employment ratios, $\Omega_{a,j,1980}$, affect changes in wages over time only through changes in gender composition. This assumption is equivalent to the common-trends assumption in a difference-in-difference framework. It assumes that, had the gender composition of occupations not changed, occupations that were gender-balanced, conditional on education, in 1980 would have experienced similar changes in wages as would occupations that were male-dominated or female-dominated.

This common-trends assumption will hold if base-year gender ratios reflect either idiosyncratic characteristics of occupations or persistent characteristics that do not have a changing relationship with wage. For instance, if women are more likely than men to work in caring occupations (Hirsch & Manzella, 2015), this might result in higher fractions of highly educated women choosing medical professions over engineering professions across the entire sample period. Likewise, idiosyncratic aspects of an occupation's culture or history can have longstanding effects on its gender composition.

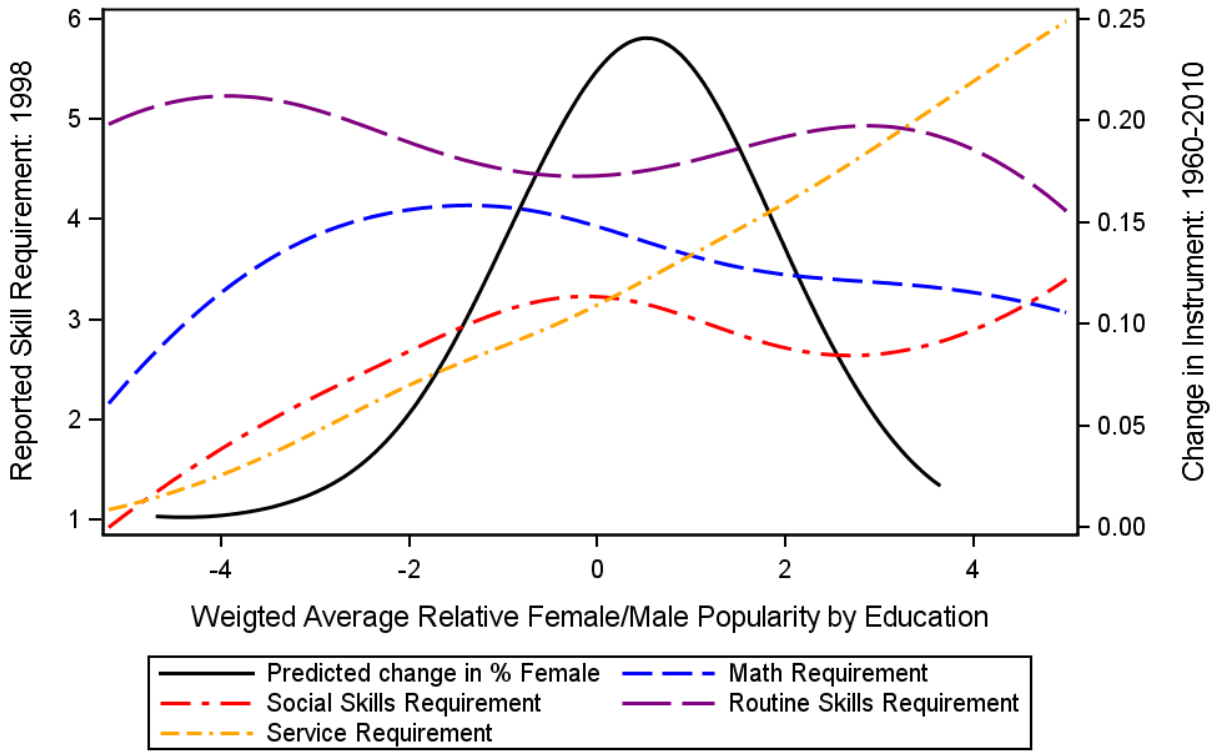
For instance, while the training, skills and work environments of dentists and veterinarians are very similar, dentistry was a far less popular occupation for female advanced degree holders than was veterinary medicine in every year of the data. As a result, dentists have a very low residual predicted growth rate in percent female over the sample period, while veterinarians have a fairly high residual growth rate in percent female. While a discussion of the relatively low number of female dentists by the American Dental Association highlighted a few challenges faced by women in dentistry, such as sexism by older male professors and practice owners and a lack of ergonomic dental tools designed for smaller hands (Solana, 2016), these factors likely reflect rather than determine the profession's gender composition.

The argument for the exogeneity of base-year gender ratios could be violated in two cases. First, gender ratios in 1980 could reflect persistent differences in occupations that have changing effects on wages over time. For example, among occupations where at least 50% of workers have bachelor's degrees, math requirements are lower in occupations that were relatively common among female workers in 1980 than in occupations that were relatively common among male workers (Appendix Figure A.6). As a result, if the returns to math skills increased from 1960-2010, as suggested by Autor Levy and Murnane (2003), we would expect wages to increase in highly-educated male-dominated occupations and to decrease in highly-educated female-dominated occupations, regardless of changes in their gender compositions.

I investigate this potential source of bias by examining the relationship between potential confounders and 1980 gender ratios. I focus in particular on four skill and task indices identified by Autor Levy and Murnane (2003) and Deming (2017) as having large changes in returns over the past half-century. Using definitions from Deming (2017), I examine indices of mathematical/analytical skills, social skills, routine tasks

and service-sector tasks measured in the 1998 o*net. These skill measures are described in detail in Deming (2017). As shown in Figure 1.4, while these skill and task indices do have relationships with the relative popularity of occupations by gender in 1980, none of them have peaks or troughs among occupations with mixed base-year gender composition. As a result, effects of these skills can be distinguished from effects of gender composition. As a robustness check, I include time-varying controls for the skill composition of occupations, and find that, as shown in Appendix Table A.5, these controls have a very small effect on the estimated relationship between fraction female and wage.

Figure 1.5: Relationship Between Skill Requirements and Weighted Relative Female/Male Popularity by Occupation



Note: Weighted Average Relative Female/Male Popularity by Education represents the ratio of the percent of female workers of each education level in an occupation to the percent of male workers of the education level in an occupation. This graph shows the nonparametric relationship between each skill and this average ratio, calculated using a penalized b-spline

The argument for the exogeneity of base-year gender ratios could also be violated if gender ratios in 1980 could reflect trends in characteristics like the factor share of labor or work amenities that are reflected in wage. For instance, if some occupations have allowed steadily more flexible working hours from 1960 to 2010, those occupations would be expected to have more flexible hours, and thus more female employees in 1980 than would comparable occupations that did not experience greater flexibility in working hours. As a result, they would expect to have declining wages over the sample period, as a consequence of steadily increasing flexibility.

I address this challenge by examining the relationship between the instrument and earlier realizations of log average wage. A valid instrument should affect wage in future periods but should not affect wage in past periods. As a result, future realizations of the instrument should be unrelated to present wage once controlling for the present realization of the instrument. I test this hypothesis by including $\widetilde{f_{J,t+10}}$ *and* $\widetilde{f_{J,t+20}}$ as controls in a regression of $W_{g,j,t}$ on $\widetilde{f_{J,t}}$. As shown in Appendix Table A.6, the estimated effect of future realizations of the instrument on contemporary wage are small relative to the standard error of the effect and relative to the estimated effect of the contemporary instrument. If the instrument were correlated with trends in wages, the future value of the instrument would have a negative relationship with wage of a similar magnitude to that of the contemporaneous value of the instrument.

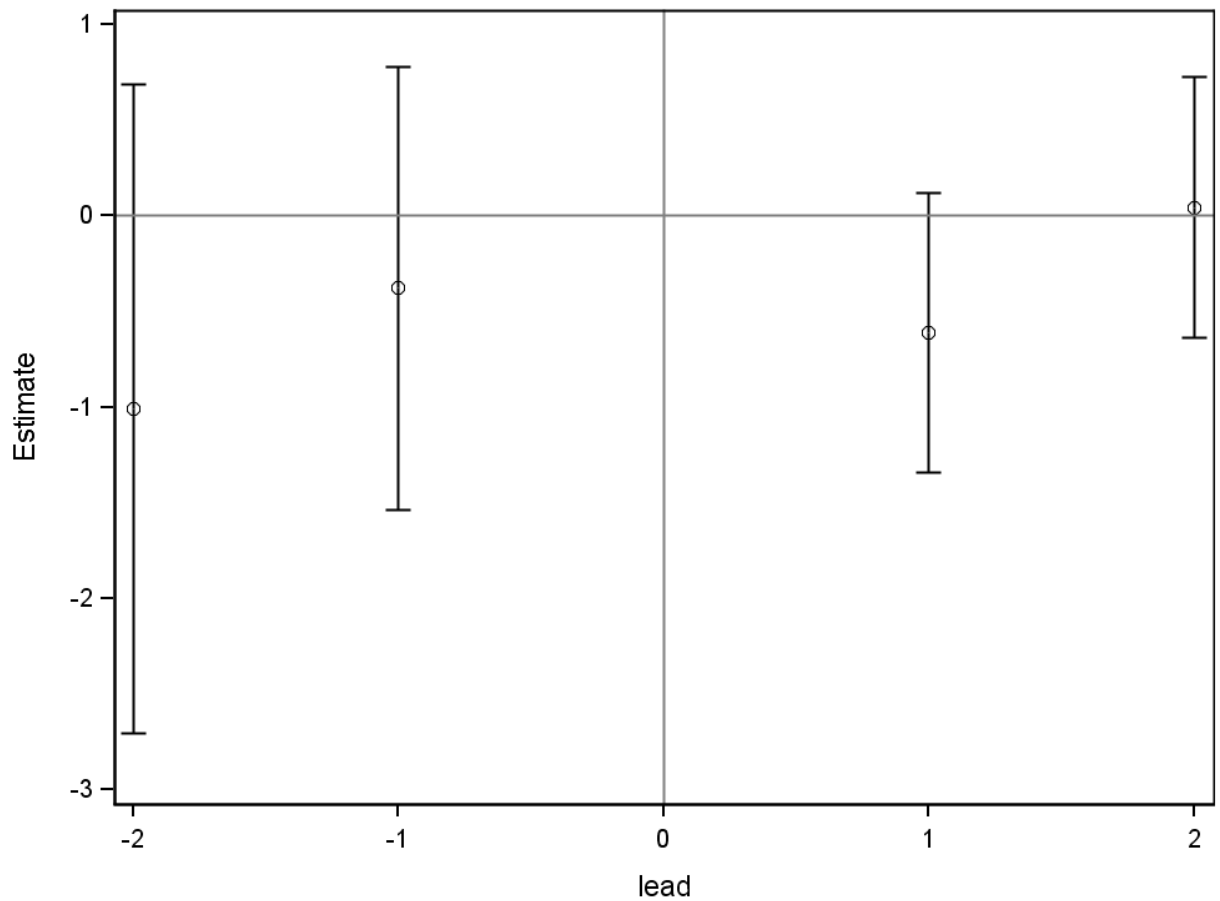
As a second, more data-intensive approach, I simulate event studies by running the following regressions, for time-lags of -20, -10, 0, 10 and 20, indexed by k:

$$W_{g,j,t+k} = \alpha + \beta_1 \widetilde{f_{J,t}} + \beta_2 \widetilde{f_{J,t+k}} + \delta X_{j,t} + \gamma_j + \sigma_t + \varepsilon_{t,j} \qquad (8)$$

The coefficient β_1 is presented in Figure 1.6 for male wage and in Figure 1.7 female wage. Because the instrument is highly autocorrelated, the event study on male wage is inconclusive. However, the event study on female wage is consistent with the

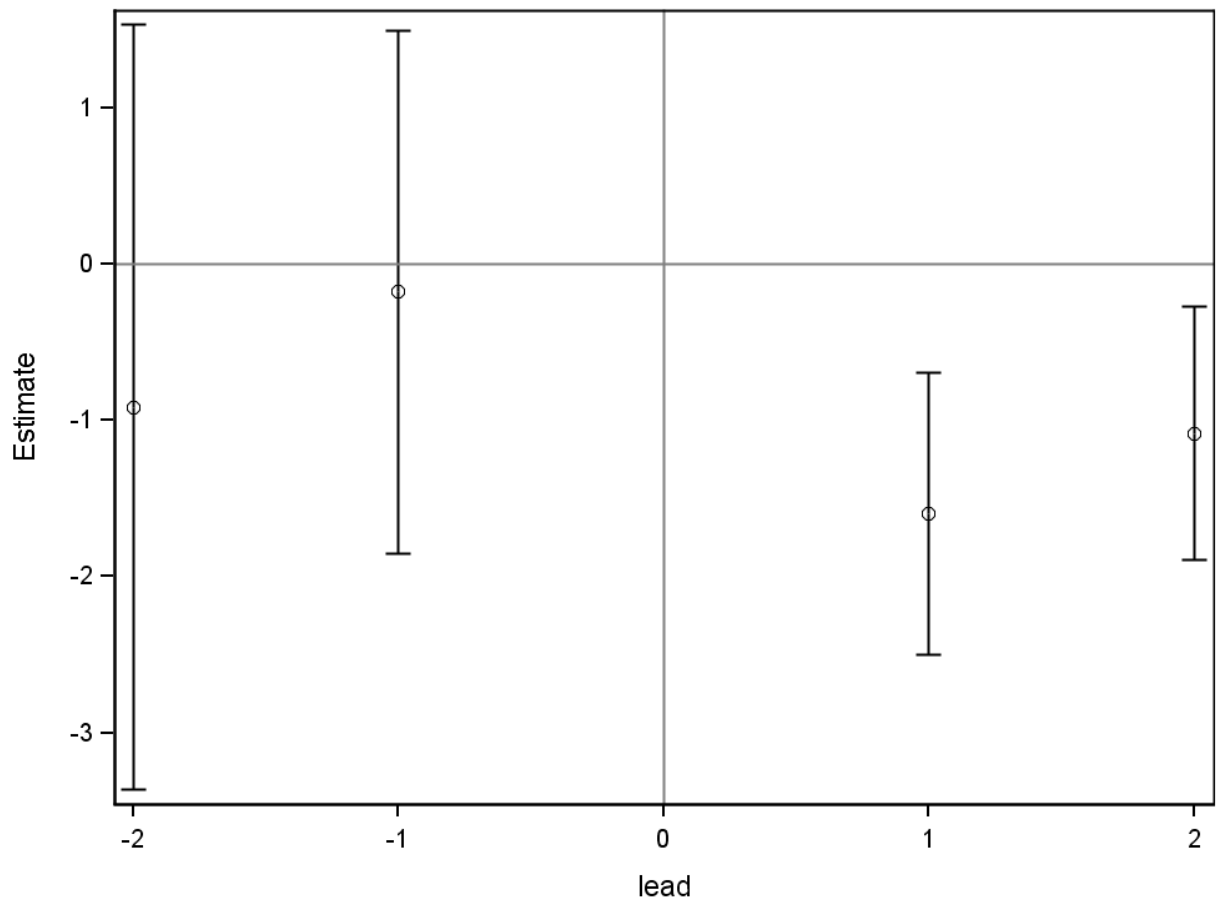
identification strategy. There is no significant relationship between future realizations of the instrument and female wage, but there is a relationship between past realizations of the instrument and female wage. Were the instrument capturing the effect of a linear time trend, the coefficients should have a linear trend, rather than a discontinuity at t=0.

Figure 1.6: Event study with Male Wage as dependent variable:



Notes: This figure presents estimated values and 95% confidence intervals of the coefficient β_1 in equation 8, estimated for female wages, with time lags of -20 years,
-10 years, 10 years, and 20 years. Each estimate reflects the effect of the instrument in year *t* on log mean male wage in year *t+k*, controlling for the instrument in year *t+k*. Standard errors are clustered at the occupation level.

Figure 1.7: Event Study with Female Wage as Dependent Variable:



Notes: This figure presents estimated values and 95% confidence intervals of the coefficient β_1 in equation 8, estimated for female wages, with time lags of -20 years,
-10 years, 10 years, and 20 years. Each estimate reflects the effect of the instrument in year *t* on log mean female wage in year *t+k*, controlling for the instrument in year *t+k*. Standard errors are clustered at the occupation level.

An alternative argument for validity of the instrument rests on exogeneity of the yearly percent female among workers of each educational attainment to dispersion in wages. The argument for exogeneity of the gender composition of education groups with respect to the relative wage of occupations depends on the assertion that trends in men's and women's education and labor supply decisions originate from sources other than the wages paid in particular occupations. Goldin, Katz and Kuziemko (2006) identify several reasons behind the increase in women's educational attainment and labor force participation, such as relative increases in girls' aptitude test scores and likelihood to take math and science courses in high school, increasing fractions of girls expecting to work as adults, and delayed marriage. Currie and Morretti (2003) further note that women increased college attendance with the greater availability of coeducational schools, which increased in numbers considerably in the 1960's and 1970's (Goldin & Katz, 2011). Several papers have found that boys are more negatively affected than girls by adverse childhood environments, with Bertrand and Pan (2013) finding that negative home environment leads to greater behavioral problems for boys than for girls, and Autor et al (2016) finding that boys benefit more from high quality schools than do girls. This work suggests that the rise in single parent households and increases in economic inequality may contribute to the increase in women's relative education by rewarding girls' greater resilience. Likewise, the increase in women's relative labor force participation has been linked to several broad changes in the labor market, such as the elimination of policy barriers keeping married women out of teaching and clerical work in the 1950's (Goldin C. , 1988), changing expectations and social norms around women's labor force participation in the 1960's and 1970s, delayed marriage and child bearing, and a shift in women's identities that increased the value of career success (Goldin, Katz, & Kuziemko, 2006).

The exogeneity of the gender composition of educational attainment groups faces two challenges. First, exogeneity would be compromised if men or women choose whether or not to attend college or graduate school by looking at the earnings of particular occupations. For example, if higher expected earnings for doctors led women to increase their graduate school attendance relative to men, this would result in covariance between the residual earnings of doctors and the gender composition of graduate school. Because $f_{a,t}^{\bar{J}}$ leaves out those in the aggregate occupation that includes j, women would need to enter graduate school on the basis of doctor's wages, but then not enter the medical field. This assumption would hold, however, if the gender composition of college graduates or postbaccalaureate graduates was associated with general increases or decreases in wages across occupations, because these general increases and decreases would be absorbed by the year fixed-effect. One piece of evidence against this concern is that controls for $W_{g,j,t}$ have no appreciable effect on the first-stage relationship between $\widetilde{f_{J,t}}$ and $f_{t,j}$ nor on the reduced-form relationship between $\widetilde{f_{J,t}}$ and $W_{g,j,t+10}$. If women and men choose levels of educational attainment based on expected future wages in particular occupations, controls for current period wage should matter. That they do not is evidence against this source of bias.

A more serious challenge to the exogeneity of gender composition of education groups with respect to relative wage is that the trends in the gender composition of education groups have been fairly smooth and linear over the past fifty years, particularly for bachelors' degree holders and advanced degree holders. As a result, the residual variation in the instrument after controlling for occupation linear time trends is insufficient for analysis. Identification from time trends in the gender composition of education groups could thus capture the effect of other trends that differentially affect the wages of occupations in the same time period. This concern is alleviated by the fact

that future values of the instrument have a weak relationship with wage when controlling for current values of the instrument (Appendix Table A.6).

Section 1.5: Results

1.5.1: OLS Regression

Table 1.2 shows estimates of regression equations 2 and 3 on the decennial census from 1960 to 2010. Results are shown for regressions using controls for education and age composition of each occupation, as described in Section 1.4.1, as well as for regressions that add induced labor supply and the time-varying effects of 1980 education composition as controls, as described in Section 1.4.2.2.

Table 1.2: OLS Estimates

	Log Female Wage				Log Male Wage			
	(t)	(t)	(t+10)	(t+10)	(t)	(t)	(t+10)	(t+10)
Fraction Female	-0.04	-0.10	0.02	-0.05	-0.01	-0.06	0.11	0.06
	(0.08)	(0.09)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Female Wage			0.03	0.02			0.09***	0.08**
			(0.07)	(0.07)			(0.04)	(0.03)
Male Wage			0.25***	0.26***			0.14**	0.15**
			(0.08)	(0.08)			(0.06)	(0.07)
Labor Supply		-0.02		-0.10		0.00		-0.01
		(0.13)		(0.09)		(0.05)		(0.05)
Base-year Controls		X				X		X
Sample Size	1817	1817	1457	1457	1817	1817	1467	1467

*Note: Each column reports results from an estimate of equation (1) in the paper, with Log Mean Wage for workers over the age of 45 as the dependent variable, estimated for males and females. The unit of observation is the occupation X year. Estimates control for age, education, and occupation and year fixed-effects as described in the text. Base-Year Controls are the share of the workforce in each education category in 1980, interacted with year. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

These regression estimates suggest that there is a weak relationship or no relationship between changes in the gender composition of an occupation and the average wage of that occupation. However, as discussed in Section 1.4.1.1, these results cannot be interpreted as a causal effect of gender composition on wages due to a multitude of non-causal factors that could explain the relationship between gender composition and future wage. In particular, the OLS results indicate that the rise in women's education and labor force attachment over the sample period is a significant source of positive bias in the relationship between changes in the gender composition of an occupation and changes in the wage of an occupation. In addition, measurement error is likely attenuating the OLS relationship between fraction female and wage—

because most occupations have experienced relatively minor changes in gender composition during the sample period, particularly after controlling for year fixed-effects, residual variation is heavily influenced by sampling error.

This finding contrasts with a few papers in sociology that have found large, negative relationships between the gender composition of an occupation and future wages in that occupation using similar panel designs (Catanzarite, 2003), (Levanon & Allison, 2009)[9], but is in line with panel estimates using the CPS (England, Allison, & Wu, 2007). One explanation for the wide range of estimates in this literature comes from Tam (1997), who demonstrates that the presence of multiple endogenous relationships between gender composition and wage leads panel estimates to be highly sensitive to specification.

1.5.2: Instrumental Variables Estimates

Table 1.3 shows the results of the first-stage regression for the instrument, both for the full sample—of all occupations and years included in the definition of the instrument, and the lagged sample—of occupations for which $W_{g,j,t+10}$ is defined. The instrument $\widetilde{f_{J,t}}$ is a strong predictor of gender composition in both samples, with an f-statistic of 38 in the full sample and 32 in the lagged sample. The inclusion of wage controls $W_{f,t}$ and $W_{m,t}$ have no effect on first-stage estimates, while the inclusion of induced labor supply control $\widetilde{l_{J,t}}$ substantially increases the first-stage coefficient on the instrument. The negative effect on $\widetilde{l_{J,t}}$, and the positive effect of the inclusion of $\widetilde{l_{J,t}}$ on the estimated effect

[9] Levanon, England and Allison (2009) uses a similar data source and panel framework as is used in this paper. Their results differ from the OLS estimates in this paper because their occupation definitions and functional form assumptions differ from this paper. In addition, in order to account for Nikkel bias Levanon, England and Allison fix the covariance between the residual and future values of the dependent variable to zero.

on $\widetilde{f_{J,t}}$ shows that male-dominated occupations had greater entry of female workers than implied by the instrument. As a result, previously female-dominated occupations have a lower percent female in later years than is implied by the instrument, while previously male-dominated occupations have a higher percent female in later years than is implied by the instrument.

Table 1.3: First-Stage Regression

	Fraction Female (Full sample)			Fraction Female (Lagged Sample)		
parameter	(3)	(2)	(1)	(6)	(5)	(4)
Instrument	0.51*** (0.11)	0.80*** (0.12)	0.80*** (0.13)	0.57*** (0.13)	0.87*** (0.14)	0.87*** (0.15)
Labor Supply		-0.19*** (0.04)	-0.19*** (0.04)		-0.19*** (0.05)	-0.19*** (0.05)
Log Male Wage (t)			0.01 (0.03)			0.01 (0.03)
Log Female Wage (t)			-0.01 (0.01)			-0.02 (0.01)
Number of Observations	1817	1817	1817	1457	1457	1457
Number of Occuaptions	380	380	380	379	379	379

*Note: Each column reports results from an estimate of equation (5) in the paper, with Log fraction of workers aged 22-65 who are female as the dependent variable, for occupations that have measured wages for male or female workers. The unit of observation is the occupation X year. Estimates control for age, education and 1980 education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

The two-stage least squares estimates shown in Table 1.4 provide evidence that the fraction of workers who are female has large, negative effects on wages for males and females, both contemporaneously and over the following ten years. Estimated simultaneous effects are slightly higher for male than for female workers, with a 10 percentage-point increase in the fraction of female workers leading to a 7.7% decrease in average male wage and a 6.1% decrease in average female wage. Lagged effects are larger for female than for male workers, with a 10 percentage-point increase in the

fraction of female workers leading to an 11.4% decrease in average female wage and an 8.1% decrease in average male wage. However, neither of these differences are statistically significant.

Table 1.4: Two-Stage Least Squares Estimates of percent female on Log Mean Wages

	Females					Males				
	Log Wage (t)		Log Wage (t+10)			Log Wage (t)		Log Wage (t+10)		
2SLS:	(1)	(2)	(3)	(4)	(6)	(7)	(8)	(9)	(10)	(12)
Fraction Female	-0.95*	-0.61**	-2.01***	-1.31***	-1.14***	-0.92***	-0.77***	-1.18***	-0.95***	-0.81***
	(0.62)	(0.36)	(0.71)	(0.4)	(0.39)	(0.34)	(0.26)	(0.4)	(0.3)	(0.27)
Labor Supply		-0.11		-0.26 **	-0.25 **		-0.05		-0.08	-0.07
		(0.11)		(0.11)	(0.11)		(0.05)		(0.06)	(0.06)
Log Male Wage (t)					0.23 **					0.13 **
					(0.1)					(0.08)
Log Female Wage (t)					-0.01					0.06 **
					(0.07)					(0.04)
First-Stage:										
Fraction Female	0.51 ***	0.8 ***	0.57 ***	0.87 ***	0.87 ***	0.51 ***	0.8 ***	0.56 ***	0.85 ***	0.85 ***
	(0.11)	(0.12)	(0.13)	(0.14)	(0.15)	(0.11)	(0.12)	(0.13)	(0.15)	(0.15)
Reduced-Form:										
Fraction Female	-0.48 *	-0.49 *	-1.15 ***	-1.14 ***	-0.99 ***	-0.47 ***	-0.62 ***	-0.66 ***	-0.81 ***	-0.68 ***
	(0.28)	(0.28)	(0.27)	(0.28)	(0.28)	(0.15)	(0.19)	(0.16)	(0.21)	(0.2)
Sample Size	1816	1816	1456	1456	1456	1816	1816	1466	1466	1466

*Note: Each column reports results from an estimate of equations (14) and (15) in the paper, with Log Mean Wage for workers over the age of 45 as the dependent variable, estimated for males and females. The unit of observation is the occupation X year. Estimates control for age, education and 1980 education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

There are a few notable characteristics of these findings that deserve some discussion. First, while the greater magnitude of lagged effects relative to simultaneous effects are not statistically significant, they could be a result of the fact that the proposed amenity and prestige mechanisms may take time to be fully realized. With respect to amenities, changes in the organization of occupations to allow flexible working hours, less competition, and other amenities of greater average value to female than to male workers may occur primarily through the formation of new firms. If so, responses to a change in gender composition may take time to occur. With respect to prestige, if changes in the prestige and advancement opportunities in an occupation lead the occupation to attract less ambitious workers, that effect will grow over time as workers change jobs and enter or exit the occupation.

Second, the lagged effect of fraction female on female wage may be larger than the lagged effect on male wage for two reasons. First, because occupations are not homogeneous, female workers may work in sub-occupations that see a larger average increase in percent female than do the sub-occupations of male workers. Estimates calculated using occupations aggregated to the sub-header level (Appendix Table A.10, described in Section 1.5.3) support this hypothesis—estimates using aggregated occupations show larger differences between the effect on male wage and on female wage than do estimates using disaggregated occupations. Second, male and female workers may be complements in production (Giorgi, Paccagnella, & Pellizzari, 2013). In this case, the presence of a large number of female workers would increase the productivity of male workers, mitigating negative effects on wages. This might be the case if, for instance, male workers are more likely to hold managerial roles than are female workers.

Finally, these results do not appear to be highly sensitive to the inclusion of wage and labor supply controls. Controlling for current-period wage does not substantially

alter the estimated effect on either male or female wage—the increase in estimated effect is less than half a standard error for males and females in the absence of a control for current-period wages. This provides evidence that these results are not driven by correlated trends in occupation wages and the gender composition of education groups. The small difference in estimated effect also reduces concern for Nikkel Bias arising out of the inclusion of a lagged dependent variable in a panel regression (Kiviet, 1995). Likewise, controlling separately for implied male and female labor supply, $\widetilde{l_{f,J,t}}$ and $\widetilde{l_{m,J,t}}$, rather than controlling for overall implied labor supply $\widetilde{l_{J,t}}$ causes a reduction in the estimated effect on male and female wage of less than a standard error.

In order to capture long-term effects of gender composition, I examine the effect of percent female in time *t* on log mean wage at time *t+20* for men and women, shown in Table 1.5. I find that while wages remain substantially lower for males and females, the effects are smaller 20 years after the change than they are 10 years after the change. A 10 percentage point increase in female share in time *t* lead to a 8.8% decrease in average female wage by time *t+20* and a 5.1% decrease in average male wage by time *t+20*. This reduced negative effect of gender composition on wage could result from firms learning how to operate efficiently with forms of organization that allow for more flexible hours and less competitive work environments. It could also reflect fade-out of negative effects of fraction female on prestige.

Table 1.5: Two-Stage Least-Squares Estimates of percent female on Wage, 20-year Lag

2SLS:	Female Wage (t+20)			Male Wage (t+20)		
	(1)	(2)	(3)	(4)	(5)	(6)
Fraction Female	-1.03 **	-0.76 **	-0.88 **	-0.52 *	-0.54 **	-0.51 *
	(0.53)	(0.37)	(0.38)	(0.38)	(0.32)	(0.33)
Labor Supply		-0.13	-0.11		0.01	0.01
		(0.11)	(0.11)		(0.08)	(0.07)
Log Male Wage (t)			0.02			-0.09
			(0.08)			(0.09)
Log Female Wage (t)			0.02			-0.09
			(0.08)			(0.09)
First-Stage:						
Fraction Female	0.66 ***	0.82 ***	0.81 ***	0.6 ***	0.77 ***	0.74 ***
	(0.16)	(0.16)	(0.16)	(0.16)	(0.17)	(0.17)
Reduced-Form:						
Fraction Female	-0.68 **	-0.62 **	-0.71 ***	-0.31	-0.42 *	-0.38 *
	(0.27)	(0.27)	(0.26)	(0.2)	(0.23)	(0.23)
Sample Size	1097	1097	1097	1116	1116	1116

*Note: Each column reports results from an estimate of equations (5), (6) and (7) in the paper, with Log Mean Wage for workers over the age of 45 as the dependent variable, estimated for males and females. The unit of observation is the occupation X year. Estimates control for age, education and 1980 education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

One puzzle in interpreting these results is why the estimated causal effect of fraction female on wage is so large. While female-dominated occupations pay less than do male-dominated occupations, the cross-sectional relationship between log average wage and fraction female is smaller than the estimated causal effect of fraction female on average wage. When controlling for the age and educational composition of occupations, the cross-sectional effect of a 10 percentage point increase of percent female in 2010 was 4.2% for female wage and 4.8% on male wage (Appendix Table A.1). Thus, extrapolating the local average treatment effect estimated in this paper to

differences in cross-sectional gender compositions would imply that the causal impact of gender composition explains or more than explains differences in pay between male-dominated and female-dominated occupations. However, there are several explanations for this finding that allow intrinsic characteristics of occupations (any characteristic not affected by the gender composition of the occupation) to have some effect on wages.

First, because female workers are more educated on average than are male workers, cross-sectional differences in wage between male and female workers are greater when including education controls. Because the education controls estimated from the decennial census are fairly broad (for instance, aggregating masters, professional and doctoral degrees), it is possible that the cross-sectional differences in wage would be larger if including a richer set of education controls. Because the estimations of the causal impact of fraction female on wage include occupation fixed-effects, they effectively control for arbitrarily detailed differences in education by occupation (averaged across time-periods). As a result, the estimated causal effect of changes in gender composition on wages may be a smaller fraction of a more fine-grained measure of the cross-sectional relationship between the gender composition and wage of an occupation.

Second, I find evidence that the effects of a change in gender composition are not entirely persistent. The estimated effect of a change in gender composition at time t is larger on wage in time $t+10$ then on wage at time $t+20$, with effects on female wage falling from 11% to 8%, and effects on male wage falling from 8% to 5%. This fade-out is theoretically reasonable—for instance, if occupations begin to reorganize in order to reduce the returns to long hours, the reorganization may be costlier to productivity in the short-run than in the long-run. As a result, the long-run equilibrium effect of fraction female on wage could be quite a bit smaller than the estimated 20-year effect.

In this case, the negative wage effects of a high percent female in occupations that have been female-dominated for a long time, like nursing, may be very small.

1.5.3: Robustness Checks

1.5.2.1: Alternative Instruments:

I define two alternative instruments that may be less exposed to confounding time trends than is the primary instrument. While these instruments are too weak to provide consistent, unbiased estimates, they produce results that are in line with the results from the main specification.

First, I calculate an alternative instrument that holds fixed the college major decisions, rather than the occupation decisions, of workers. While college major choices are influenced by expected future occupation, the college major decision is made at a degree of remove from the labor market, and at a time when students have less knowledge about the earnings and job amenities of various occupations. As a result, gender differences by major choice may be less likely to reflect changes over time in the characteristics of the occupations they feed into than do gender differences in the occupations themselves. As shown in Table 1.6, men and women choose different college majors, conditional on their level of educational attainment. Because of these persistent differences in the college major choices of women and men, the effect of women's increased educational attainment on gender composition is greater for occupations that predominantly hire from gender-balanced college majors, like Biological/Life Science rather than from male or female-dominated majors, like Education or Physical Science/Engineering.

Table 1.6: Male and Female College Major Choice, 2009-2014

	College Only		Master's or Higher	
Major Field of Study	%of Males	%of Females	%of Males	%of Females
Medical and Health Services	2.4	11.0	3.3	8.9
Education	3.9	12.0	6.2	17.0
Social Work/Interdisciplinary/Psychology	5.8	12.0	6.8	16.0
Humanities/Communications/Law	15.0	20.0	12.0	16.0
Biological/Life Sciences	5.9	5.1	11.0	9.8
Business	27.0	23.0	16.0	12.0
Social Science	10.0	7.5	13.0	10.0
Construction/Manufacturing/Criminal Justice	4.7	2.1	1.6	1.1
Math/Computer Science	8.3	2.9	7.3	3.2
Physical Sciences/Engineering	17.0	4.7	23.0	7.0

*Note: Calculations from the 2009-2014 American Community Survey. Sample includes all respondents who include any major and any education. Includes only first major listed.

This instrument estimates the fraction of available workers in each occupation who would be female if the occupation hired the same proportion of workers from each education level and college major as it did in 2010, if male and female students had the same propensity to choose each college major as they did in 2010, but if the fraction female of each educational attainment group varied over time. I define the instrument as:

$$\widetilde{f_{t,J}}^{M} = \sum_{A,M} \gamma^{j}_{am,2010} * \left(\frac{\omega^{F}_{am,2010} * f^{\bar{J}}_{a,t}}{\omega^{F}_{am,2010} * f^{\bar{J}}_{a,t} + \omega^{M}_{am,2010} * \left(1 - f^{\bar{J}}_{a,t}\right)} \right) \quad (10)$$

Where $\omega^{F}_{am,2010}$ is the fraction of females with educational attainment a choosing major m in 2010, $\omega^{M}_{am,2010}$ is the fraction of males with educational attainment a choosing major m in 2010, $f^{\bar{J}}_{a,t}$ is the female fraction of workers with educational attainment a in year t, excluding those employed in occupation j, and $\gamma^{j}_{am,2010}$ is the fraction of workers in occupation j with education level a and major m, in 2010. Because variation in this

instrument depends on college major, which is available only for college graduates, I include only occupations where at least 25% of workers had a college degree or higher in 2010 in estimates using this instrument.

I construct a second robustness instrument by calculating the instrument defined in equation 4 using 1970 as a base year, rather than 1980. I exclude 1960 data from the estimates using this instrument, so that 1970 is both the base year and the initial year. I use 1970 as a base year, rather than 1960, because the female fraction of most professional occupations in 1960 was too low to generate reasonable estimates. This instrument is estimated as:

$$\widetilde{f_{t,J}}^{1970} = \sum_A \gamma_{a,1970}^{j} * \left(\frac{\omega_{aj,1970}^{F} * f_{a,t}^{\bar{J}}}{\omega_{aj,1970}^{F} * f_{a,t}^{\bar{J}} + \omega_{aj,1970}^{M} * \left(1 - f_{a,t}^{\bar{J}}\right)} \right) \tag{9}$$

I construct this instrument to address the possibility that the main result is confounded with trends in characteristics of occupations that are related both to wages and to the relative likelihood that males and females work in an occupation. For instance, suppose that social science and therapeutic occupations experienced decreasing returns to long hours, while returns to long hours remained high in physical science and engineering occupations. This could lead educated women to choose social science and therapeutic occupations at increasing rates relative to physical science and engineering occupations, leading to a correlation between the conditional gender ratio of occupations in 1980 and the trend in returns to hours over the sample period.

Several pieces of evidence are inconsistent with such a trend being present through the full sample period. Because such a trend would result in the shift-share instrument not fully accounting for changes in occupational sorting between men and women, it would bias a shift-share instrument in the direction of OLS results, which do

not account at all for changes in sorting. Instead, I find instrumental variables estimates that are far larger than the OLS estimates. Additionally, as shown in Appendix Table A.6, future values of the instrument are not correlated to wage, controlling for the current value of the instrument. However, such a trend may be present for some part of the sample period. In particular, changes to occupations from computerization or automation may only begin to affect gender sorting and wages from 1980 onward. A 1970 base year eliminates trends that only begin to affect sorting by gender in 1980.

These two alternative instruments are correlated with the primary instrument, but the residual variation in the alternative instruments differs substantially from the residual variation in the primary instrument. Appendix Table A.3 shows the correlations between the residual of each instrument after controlling for occupation and year fixed effects, current-period education and age, time-varying base-year education and induced labor supply.

Table 1.7 shows the estimated effect of fraction female on wage, using these robustness instruments. These estimates are run on restricted samples. Estimates using the college major instrument, shown in columns 1, 2, 5, and 6 of Table 1.7, are run only for occupations where more than 25% of workers were college graduates in 2010, as well as occupations that were not included in the 2010 occupation codes, and do not include an induced labor supply control. Estimates using a 1970 base year, shown in columns 3, 4, 7, and 8 of Table 1.7, exclude 1960, and use an induced labor supply control, $\widetilde{l_{J,t}}^{1970}$, that is defined using a 1970 base year.

Table 1.7: Robustness Results for Two-Stage Least-Squares

	Log Female Wage				Log Male Wage			
	Major Instrument		1970 Base Year		Major Instrument		1970 Base Year	
	(t)	(t+10)	(t)	(t+10)	(t)	(t+10)	(t)	(t+10)
2SLS:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction Female	-2.15**	-2.42*	-0.62	-1.42**	0.34	-0.14	-1.42***	-0.93**
	(1.18)	(1.62)	(0.62)	(0.62)	(0.58)	(0.54)	(0.58)	(0.44)
Labor Supply (1970)			-0.21**	-0.44***			-0.22**	-0.14*
			(0.12)	(0.15)			(0.09)	(0.09)
Log Male Wage		0.09		0.26***		0.18***		0.12
		(0.18)		(0.11)		(0.07)		(0.09)
Log Female Wage		0.07		-0.17**		0.18***		0.04
		(0.14)		(0.07)		(0.06)		(0.05)
First-Stage:								
Fraction Female	0.91***	0.8**	0.61***	0.8***	0.91***	0.82**	0.61***	0.79***
	(0.32)	(0.36)	(0.14)	(0.16)	(0.32)	(0.36)	(0.14)	(0.16)
Reduced-Form:								
Fraction Female	-1.95***	-1.94**	-0.38	-1.14***	0.31	-0.12	-0.87***	-0.74**
	(0.73)	(0.91)	(0.37)	(0.43)	(0.52)	(0.43)	(0.29)	(0.32)
Sample Size	827	669	1348	1059	827	670	1348	1071

*Note: Each column reports results from an estimate of equations (14) and (15) in the paper, with Log Mean Wage for workers over the age of 45 as the dependent variable, estimated for males and females. This table provides estimated effects using two alternative instrument definitions. "Major Instrument," the instrument in specifications 1, 2, 5, and 6, constructs the instrument by holding constant the fraction of workers from each college major entering an occupation, with a base year of 2010. Analysis using "Major Instrument" drops all occupations where fewer than 25% of workers held a bachelor's degree or higher in 2010. "1970 Base Year," the instrument in columns 3, 4, 7, and 8 uses 1970, rather than 1980 as a base year. These regressions drop the year 1960 from the analysis. The unit of observation is the occupation X year. Estimates control for age, education and base-year education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

These robustness checks produce fairly noisy estimates, but the results are broadly consistent with the results of the primary analysis. Estimates using the college major instrument indicate that a 10 percentage point increase in fraction female leads to a 22% decline in contemporaneous female wages and a 24% decline in female wages over 10 years. The estimated effect on male wages is noisy and insignificant, with a 10 percentage point increase in the fraction female leading to a 3% increase in contemporaneous male wage and a 1% decrease in male wage over 10 years. Estimates using a 1970 base year are also broadly consistent with those using the primary instrument. Using a 1970 base year, a 10 percentage point increase in fraction female leads to an estimated contemporaneous 6% decline in average female wage and a 14% decline in average male wage. Over 10 years, the effect shifts to a 14% decline in average female wage and a 9% decline in average male wage. Because the standard errors on these estimates are very wide, the difference between these results and the primary results may be statistical noise. This is especially likely because estimates from each robustness instrument are not consistently larger or smaller than the primary results.

I present seven additional robustness checks, with results in Appendix Tables A.7-A.12. Each is described in turn below.

Estimate using College Graduates only (Appendix Table A.7)

I estimate the effect of changes in gender composition on the wages of college graduates as an additional test for bias resulting from changing returns to education. Because college-educated workers saw increasing relative wages over this period, changing returns to education could be confounded with changes in the gender composition of education groups. As show in Table A.7, the estimated effect of a change in fraction female on log male and female wage is similar for college graduates and for all workers. An increase in the fraction female of 10 percentage points leading to a 6%

decrease in female wages and a 10% decrease in male wages contemporaneously, and a 13% decrease in female wages and a 9% decrease in male wages over 10 years.

Estimate wage effect on "Newcomer" workers aged 20-35 (Appendix Table A.8)

I examine the wages of young workers in order to determine whether the estimated effects of fraction female on log average wage reflect complementarities between young and old workers. In particular, a negative effect on the wage of older workers from an increased fraction female among younger workers could be caused by young male workers enhancing the productivity of older workers. Because the estimated effect of fraction female on the average wage of young workers (Table A.8) is similar to the effect on old workers, this does not appear to be the case. One important exception is that the estimated lagged effect for young male workers is less negative than is the estimated lagged effect for older male workers, with a 10 percentage point increase in the fraction female leading to a 3% decline in young male wage and an 8% decline in older male wage over 10 years. This may be a consequence of competition between young workers—young male workers may be more likely to be promoted when a larger fraction of the young workforce is female, either due to gender bias or due to differences in preferences. Additional evidence for the effect of competition comes from the greater estimated labor supply elasticity for young male and female workers than for older workers, with a 10% increase in the labor supply of young workers leading to a 3% decrease in the wages of young female workers and a 1.7% decrease in the wages of young male workers, measured contemporaneously. The same increase in labor supply leads to a 1.1% decrease in the wages of older female workers and a 0.5% decrease in the wages of older male workers.

Estimate results without labor supply controls and without time-varying base-year education controls (Appendix Table A.9).

I examine the effect of fraction female on log wage, removing labor supply and time-varying education controls from the analysis. Time-varying education controls are included in the main analysis because the instrument is positively correlated with the education level of the occupation in the base year. While time *t* education controls are included in all models, it is possible that occupations that had higher measured education requirements in 1980 also have higher unmeasured education requirements in all years, conditional on measured education requirements in time *t*. However, because the inclusion of time-varying education controls absorbs a great deal of variation in the instrument, it is worthwhile to see whether the results are robust to the exclusion of this control. While the exclusion of 1980 base-year controls reduces the magnitude of the estimated effect of fraction female on average wage, particularly for male wages, estimated effects remain substantial and negative. A reduction in the estimated effect of fraction female on average wages is consistent with increasing returns to education.

Estimate using aggregate occupation definitions (Appendix Table A.10)

Changes in the classification of occupations over the period 1960 to 2010 add noise to the estimated effect of fraction female on average wage. These reclassifications could also bias the estimated relationship if high-wage jobs are likely to be classified differently than similar, predominately female jobs. To address this, I estimate the relationship between fraction female and wage using 1990 occupation definitions, aggregated to the sub-header level. The three-digit Standard Occupation Codes used in the decennial census are published with codes grouped under thematic sub-headings. For example, Physicians, Dentists, Veterinarians, Optometrists, Podiatrists, and Health

diagnosing practitioners, n.e.c. are all included in the sub-heading "Health Diagnosing Occupations." I use sub-heading aggregation, rather than aggregating Standard Occupation Codes to the two-digit level, because occupations grouped by sub-heading are more consistent with each other than are occupations with the same first and second digit, and because these occupations were intentionally categorized together in the construction of the occupation codes.

As shown in Appendix Table A.10, a 10 percentage point increase in the fraction female of an aggregate occupation leads to an estimated 8% decline in average female wage and a 6% decline in average male wage, measured contemporaneously. Over 10 years, the effect grows to a 14% decline in average female wage and a 6% decline in log male wage. While these results are broadly consistent with the estimated effects on 3-digit occupations, effects on male wages are smaller than on female wages. This may be a consequence of male workers having smaller changes in their specific occupations than is implied by the gender composition of aggregated occupations.

Estimate including control for employment as a share of the total workforce (Appendix Table A.11)

I include controls for the log of current period employment (measured as the fraction of all workers who work in occupation j at time t). I do this to address concerns that the labor supply index, defined in equation 4, does not fully account for changes in labor supply of occupations induced by women's increased education and workforce participation. A possibility of particular concern is that the occupations with the greatest increase in fraction female also saw the greatest actual increase in labor supply. This would occur if, among occupations that were open to women, women chose gender-balanced occupations in greater numbers as their share of the workforce increased. Despite these concerns, I exclude actual employment from my main specification for two

reasons. First, because changes in selection into an occupation are a causal channel, employment controls may absorb some of the causal impact of gender composition on wage. Second, because occupation reclassifications have a greater effect on employment than on wage, employment data from the census is noisy.[10] As shown in Table A.9, employment controls do not change the estimated effect of fraction female on wage.

Estimate Long-Difference Results (Appendix Table A.12)

As a final test of the effect of gender composition on wage, I estimate the relationship between expected change in percent female over the period from 1960 to 2010 and the change in the average wage of male and female workers in the same period. The long difference approach is valuable for two reasons. First, it does not take a stance on the timing of the effect of fraction female on log wage. Second, it accepts the relative linearity of trends in the gender composition of education groups by treating the entire period as a single shock.

I estimate long-difference models using first-difference estimation for occupations that were included in the 1960 and 2010 census. I calculate Δf_j, $\Delta \widetilde{f_j}$, $\Delta \widetilde{l_{j,t}}$ and $\Delta W_{g,j}$ as the difference between the 2010 and 1960 values of fraction female, instrumented fraction female, induced labor supply and wage respectively. I then estimate the following two-stage least-squares equations, with results shown in Appendix Table A.12:

First-Stage:

$$\Delta f_j = \beta_1 * \Delta \widetilde{f_j} + \beta_2 * \Delta \widetilde{l_j} + \beta_3 \Delta X + \varepsilon_j \tag{11}$$

Reduced-Form:

$$\Delta W_{g,j} = \beta_1 * \Delta \widetilde{f_j} + \beta_2 * \Delta \widetilde{l_j} + \beta_3 \Delta X + \varepsilon_j \tag{12}$$

[10] Employment is more affected by occupation reclassifications than is wage or gender composition because employment is affected by all grouping or splitting of occupations, even if the grouped and split occupations are very similar. In contrast, grouping two similar occupations (that have similar gender compositions and wages) together will have a modest effect on measured average percent female and wage.

As shown in Table A.12, the results of the long-difference regression are noisy and are not statistically significant at conventional levels, but are broadly consistent with the main estimates of the effect of fraction female on wage. A 10 percentage-point increase in fraction female leads to an estimated decline over the sample period of 3.6% for men and 9.1% for women. These estimates are in line with the 20-year estimated effects of 8% for women and 5% for men.

Section 1.6: Mechanisms

As discussed in Section 1.2, while there are several mechanisms by which the gender composition of an occupation can affect wages paid in that occupation, the two sets of mechanisms most likely responsible for these results are changes in workplace amenities and changes in prestige. The amenities mechanism operates through differences in the preferences of male and female workers. As an occupation sees an increased percentage of female workers, the hours requirements, competitive pay structures, and other job characteristics adjust to reflect the preferences of female workers. The prestige mechanism operates through differences in the perceived abilities of male and female workers. As an occupation sees an increased percentage of female workers, male and female workers in the occupation are perceived to be of lower ability or lower quality by outside employers, causing reduced advancement opportunities and making the occupation less attractive to high ability/high ambition workers. These two mechanisms are not mutually exclusive, and may in fact be mutually reinforcing.

I perform two tests of the amenities mechanism. First, I examine the effect of a change in fraction female on the average number of hours worked per week by men and women. If women pay higher costs for long working hours than do men, increased

representation of women in an occupation will create incentives for firms to organize work in a way to reduce the cost of short working hours. As a result, the average number of hours worked should decrease for both male and female employees. I test this hypothesis by estimating a two-stage least-squares regression of average hours worked on fraction female. These results are presented in Table 1.8.

Table 1.8: Effect on hours worked

	Female Avg. Hours Worked		Male Avg. Hours Worked	
2SLS:	(t)	(t+10)	(t)	(t+10)
Fraction	-1.3	-3.52	-2.81	-3.11
Female	(5.25)	(4.21)	(4.12)	(3.19)
Labor	-1.66	-1.9**	-2.66***	-3.78***
Supply	(1.45)	(1.12)	(0.81)	(0.82)
		-1.63**		-0.86
Male Wage		(0.98)		(0.84)
Female		-0.11		0.06
Wage		(0.87)		(0.3)
First-Stage:				
Fraction	0.8***	0.87***	0.8***	0.85***
Female	(0.12)	(0.15)	(0.12)	(0.15)
Reduced-Form:				
Fraction	-1.03	-3.05	-2.24	-2.63
Female	(4.17)	(3.63)	(3.28)	(2.66)
Sample Size	1816	1456	1816	1466

*Note: Each column reports results from an estimate of equations (14) and (15) in the paper, with average hours worked for workers over the age of 45 as the dependent variable, estimated for males and females. The unit of observation is the occupation X year. Estimates control for age, education and 1980 education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

As shown in Table 1.8, there is a negative relationship between fraction female and hours worked, consistent with an amenities channel. The estimated effect of gender composition is similar in magnitude to the difference in average hours worked between

male and female workers. However, the difference is not statistically significant when controlling for labor supply.

I perform a second test of amenities by estimating the elasticity of earnings with respect to hours worked per week for each occupation and year, following Goldin (2014). In each census year t, I perform the following regression at the individual level:

$$logearn_i = \alpha_{j,t} + \beta_{j,t} loghours_{i,j} + \gamma_{j,t} fem_{i,j} + \delta_t X_i + \varepsilon_i \qquad (13)$$

Where $logearn_i$ gives log annual earnings for worker i, $loghours_{i,j}$ gives the log number of hours per week reported by worker i, $fem_{i,j}$ indicates whether worker i is female, and X_i gives a set of individual controls that includes log weeks worked per year, a quartic of age, and years of education. $\beta_{j,t}$measures the elasticity of earnings with respect to hours worked for full-time workers in occupation j at time t. This regression is run on a sample of full-time workers (working at least 35 hours per week and 40 weeks per year) between the ages of 25 and 65. By estimating the regression equation in each year, I create a measure of the elasticity of earnings with respect to hours at the year by occupation level. Because full-time workers who work long hours may differ from those who work short hours within an occupation, the estimated elasticity of earnings with respect to hours, $\beta_{j,t}$ is biased by selection.

I measure the effect of fraction female on returns to hours worked by regressing fraction female on $\beta_{j,t}$, using the regression specification described in equation 6. In order to account for measurement error, I weight estimates by the inverse standard error of $\beta_{j,t}$. As shown in Table 1.9, changes in fraction female is positively associated with returns to hours worked.

Table 1.9: Effect on Elasticity of Earnings with respect to hours per week

2SLS:		Elasticity of Earnings With Respect to Hours/Week (t)	(t+10)	(t+20)
	Fraction Female	0.91**	0.81*	0.42
		(0.49)	(0.51)	(0.62)
	Labor Supply	-0.13	-0.08	0.06
		(0.12)	(0.12)	(0.14)
First-Stage:				
	Instrument	0.82***	0.86***	0.82***
		(0.16)	(0.17)	(0.19)
	Dependent Mean	0.31	0.39	0.44
	Sample Size	1816	1478	1156

*Note: Dependent Variable is the estimated elasticity of earnings with respect to hours worked for full-time workers in an occupation. This elasticity is given by $\beta_{j,t}$ from equation (13). Regressions are weighted by the inverse standard error of $\beta_{j,t}$. All specifications include occupation and year fixed-effects, time-varying controls and base-year education controls. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

The effects of fraction female on returns to hours worked are modest. A 10 percentage point increase in the fraction female would increase returns to hours worked by 0.091, implying that a 10% increase in hours worked would increase earnings by 0.91% more than it would have prior to the change in gender composition. This effect

fades out over time, shrinking from an estimated contemporaneous effect of 0.91 to a 20-year effect of 0.42.

This result is consistent with a higher fraction female leading to greater demand for short hours, followed by a reorganization of firms to reduce the costs of short hours. Goldin (2014) argues that increases in the number of women in an occupation should increase the returns to long hours in the short run. If demand for the amenity of short working hours increases and the supply of short working hours is upward-sloping, the cost of short working hours should increase. At the same time, greater demand for short working hours increases returns for business models that make short working hours less costly. The decreasing effect on returns to hours worked over 20 years is consistent with this analysis—as firms learn how to reduce the cost of short working hours, the earnings differences between workers working long hours and workers working short hours shifts toward its previous level, even as more workers reduce their hours. However, it is also important to note that the elasticity of earnings with respect to hours may also be biased by selection into long hours of work. Because increases in fraction female are associated with decreases in hours worked, the workers who continue to work long hours may be higher wage than those who reduce their hours.

Finally, I test the effect of changes in the female fraction of an occupation's workforce on prestige. I measure changes in an occupation's prestige by taking the difference between the prestige of an occupation measured by the 1964 Hodge-Siegel-Rossi survey (Siegel, 1971) and the prestige of the occupation measured by the 1989 General Social Survey (Nakao & Treas, 1994). Each survey measures the prestige of an occupation by asking respondents to place each occupation on a ladder with ten rungs. Scores are then calculated as the weighted average of the score given by each respondent, and placed on a scale from 1-100.

I estimate the effect of fraction female on prestige by performing a two-stage least-squares regression of the difference in measured prestige from 1964-1989 on changes in the predicted female fraction of the occupation from 1970-1990. The results are shown in Table 1.10.

Table 1.10: Effect on Prestige

	Change in Prestige (1965-1989)		
2SLS:	(1)	(2)	(4)
Change in Fraction Female (1970-1990)	-35.31*	-12.7	-12.19
	(23.49)	(15.33)	(13.37)
Change in Labor Supply (1970-1990)	0.09	-7.07***	-9.23***
	(3.3)	(2.95)	(2.97)
1980 Education		X	X
Time-Varying Controls			X
First-Stage:			
Instrument	0.37***	0.78***	0.84***
	(0.12)	(0.14)	(0.14)
Sample Size	302	302	296

*Note: Each column reports results from a regression on the difference between occupation's Nakao-Treas prestige rating in 1989 and its Hodge-Siegel-Rossi prestige rating in 1964. In both surveys, prestige ranges from 0-100. The independent variable of interest is the difference in the instrument from 1970-1990. Analysis is at the occupation level. Time-varying controls include controls for education and age, described in the text. 1980 education controls give the education composition of the occupation in 1980. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

Change in prestige has a weak negative association with the change in the fraction of female workers in an occupation. When controlling for time-varying education effects, a 10 percentage-point increase in the fraction female of an occupation is associated with a 1.2 point decline in measured prestige. Without time-varying education controls, the effect is larger, with a 10 percentage-point increase in fraction female leading to a 3.5 point decline in measured prestige. This provides suggestive evidence that changes in the gender composition of an occupation might change the perceptions of that occupation among the public.

Section 1.7: DISCUSSION

In this paper, I find the causal evidence on the effect of gender composition of an occupation on the wages of that occupation. My findings indicate that the effect of an increase in the percent female of an occupation on the wages for men and women are large and sustained, with a 10 percentage-point increase in the female share of an occupation leading to an 8% decline in average male wage and a 6% decline in average female wage. That effect grows over ten years to an 8% decline in average male wage and a 12% decline in average female wage, and shrinks over 20 years to a 5% decline in average male wage and an 8% decline in average female wage. These effects are large, and if the effect of gender composition on wage is homogeneous would explain or more than explain the cross-sectional relationship between gender composition and wage. While these estimates are imprecise, even lower-bound estimates are large relative to the absolute difference in average pay between male-dominated and female-dominated occupations.

These findings have several implications for policy and future research. Most importantly, these findings suggest that changes to occupations in response to the

gender composition of the occupation's workforce play a crucial role in producing the observed discrepancy between the wages of female-dominated and male-dominated occupations with similar education requirements. As a result of lower pay in occupations with more women, preferences and skills more common among female workers may earn lower returns in the labor market than they would were gender a less salient factor in the labor market.

These findings also suggest that many of the observable differences in the characteristics of male and female dominated jobs may be consequences of gender composition, including differences in prestige, in returns to hours, and in ownership structures. This in turns suggests that correlations between observed characteristics of occupations that are related to gender composition and wages are likely to reflect the effect of gender composition on wage, and cannot be taken as measures of the direct effect on the observable characteristics themselves.

Other implications of this work depend on deeper exploration of the mechanisms driving the result. If the decline in wages in occupations that see rising female shares of the labor force is caused by the provision of amenities valued by women, this finding may auger well for gender integration of occupations. An amenities mechanism implies that occupations with work and pay structures that do not appear conducive to female work, such as business and management (Bertrand, Goldin, & Katz, 2010) may become more conducive to female work in the presence of a larger female workforce. Likewise, it implies that were men and women to continue to become more similar in their balancing of work and non-work responsibilities (Goldin C. , 2014), differences in the amenities provided in male and female-dominated occupations, and thus the earnings of those occupations, could also become more similar.

On the other hand, if the decline in wages in occupations with rising female shares of the labor force is caused by declines in prestige for those occupations, this finding would suggest that wage gaps may persist even in the face of continued integration of currently male-dominated occupations. A prestige mechanism implies that integrating a high-paying male-dominated occupation will likely cause declines in the wages paid in that occupation for male and female workers, due to some combination of highly skilled sorting into other occupations, demand for the occupation falling, and a reallocation of high-skill tasks away from the occupation. As a result, the entry of substantial numbers of female workers into a particular highly paid male-dominated occupation may not generate as large an increase in earnings for those workers than would otherwise be expected. Future work should look more deeply into these mechanisms by examining changes in the organization of firms and the allocation of tasks among workers in response to changes in gender composition.

Professional Interactions and Hiring Decisions: Evidence from the Federal Judiciary

Marco Battaglini Jorgen M. Harris Eleonora Patacchini,
Cornell University Cornell University Cornell University
6/21/19

ABSTRACT

We study the effect of hearing cases alongside female judicial colleagues on the probability that a Federal judge hires a female law clerk. Federal judges are assigned to cases and to judicial panels at random and have few limitations on their choices of law clerks: these two features make the Federal court system a unique environment in which to study the effect of professional interactions and beliefs in organizations. For our analysis, we constructed a unique dataset by aggregating federal case records from 2007-2017 to collect information on federal judicial panels, and by merging this data with judicial hiring information from the Judicial Yellow Books, a directory of federal judges and clerks. We find that a one standard deviation increase in the fraction of co-panelists who are female increases a judge's likelihood of hiring a female clerk by 4 percentage points.
JEL Codes: J16, J82, J110
Keywords: Economics of Gender, Discrimination, Labor Force Composition

Section 2.1: Introduction

Prejudiced and discriminatory beliefs are widespread in human relationships, and have profoundly negative consequences both for members of out-groups and for society as a whole. There is a long-standing literature in sociology documenting how these beliefs are formed, when they are maintained, and how they can be disrupted. These topics are recently attracting growing attention among economists. Several theoretical papers have developed models explaining the perpetuation of different beliefs between groups under different hypotheses, including the scenario when one group's beliefs are incorrect (Golub and Jackson (2012), Algan et al. (2015), Levy and Razin (2016) and the overview in Fang et al, 2010).

To date, however, the empirical evidence has been virtually non-existent. Some contributions can be found in the literature examining the role of exposure on beliefs in the context of neighborhood segregation and ethnic minority outcomes. For example, Boisjoly et al (2006) and Burns, Corno and La Ferrara (2015) note that living in racially diverse housing can decrease discriminatory views. However, making causal inference in these contexts is notoriously difficult because location decisions and the formation of voluntary social networks and social interactions are influenced by the beliefs and prejudices of the individuals making them. Some papers have attempted to address this problem by looking at the effect of non-voluntary social relationships, such as the gender of a child or the characteristics of assigned dorm mates, class mates, or military units (Dahl, Kotsadam and Rooth 2018), on actions and beliefs. For example, Washington (2008) shows that politicians with more daughters demonstrate more feminist voting behavior, and Glynn and Sen (2015) note that the presence of daughters causes judges to give more female-friendly rulings.

In this paper, we project presents the first causal evidence on the effects of exposure to out-groups for individual hiring decisions. We do so using the unique and highly relevant context of the Federal appellate court system. Appellate court cases are heard by panels of three judges, randomly selected from a pool of appellate court justices and district court justices. Because appellate judges do not choose the cases that come before their courts or the colleagues with whom they hear these cases, their likelihood of working with female colleagues on cases is effectively random. At the same time, appellate judges are broadly unconstrained in their decision of who to hire as a court clerk, a highly prestigious position typically filled by graduates of top law programs. As a result, changes in the likelihood of hiring a female clerk likely reflect changes in a judge's assessment of the likely ability of a female junior colleague. Because it is identified from randomly assigned interactions, this paper suffers from few challenges to causality relative to the social networks literature. Moreover, this paper is unique in the non-voluntary interactions literature because we examine high-stakes, sustained, professional interactions between peers at the elite of their occupations.

We find substantial positive effects on the likelihood to hire female clerks from exposure to a larger number of female judges. In particular, we find that a one standard deviation increase in the fraction of case interactions a judge carries out with female colleages increases the likelihood of hiring at least one female clerk in the next year by 5 percentage points. Consistent with the hypothesis that these changes in hiring reflect changes in beliefs, we find suggestive evidence that these effects are larger for male judges, for judges with few women currently on their staff, and for less experienced judges.

These findings contribute to the literature by providing novel information about the formation and dissolution of in-group prejudice in the job market, particularly in the upper reaches of the job market, where wage gaps are greatest (Blau and Kahn 2016). Because randomization of influence is difficult to come by, most prior work on this subject has either taken advantage of randomization of peers in college (Mark and Harris 2012h), in military training (Dahl, Kotsadam and Rooth 2018) or of the effectively random gender of a biological child (Washington 2008, Glynn and Sen 2015). While this literature has demonstrated that in-group prejudice is influenced by exposure to out-groups, with students assigned a different-race roommate likely to have more friends of a different race (Mark and Harris 2012), politicians more likely to vote liberal after the birth of a daughter (Washington 2008), and military cadets assigned to mixed-gender units more likely to adopt egalitarian attitudes (Dahl, Kotsadam and Rooth 2018), no previous paper had examined the effect of professional interactions on hiring. While Dahl, Kotadam and Rooth also examine the role of professional interactions on beliefs, their work focuses on military recruits at the beginning of their career, when we might expect beliefs about gender and professional ability to be most malleable. The finding that these beliefs are also malleable among middle-aged professionals with significant power to shape their professions extends this literature substantially. Other work examining judge assignment to appellate panels, such as Boyd, Epstein and Martin (2010) have examined the effect of female judges on the rulings of their colleagues, but have not looked at the effect of interacting with female colleagues on decisions made independently, such as hiring.

Section 2.2: Institutional Setting

2.2.1: Federal Appellate Courts

The federal court system in the United States is composed of three tiers: District Courts, Appellate Courts, and the Supreme Courts. All cases are initially heard in district courts, where evidence is presented, parties appear, and an initial ruling is made. Parties bringing a case in federal court are entitled to appeal decisions to Appellate Courts, which review the legal reasoning used in District Courts. Parties can request an appeal of appellate court decisions in the supreme court, but are not entitled to a review—as a result, the appellate court's decision is final in the great majority of cases.

Federal Appellate courts are organized into circuits, which are primarily organized geographically. Twelve circuits cover cases heard in specific geographical areas, while the federal circuit court hears cases involving patents, international trade, money claims against the Federal Government, and other particular subject areas originating anywhere in the United States. Each appellate court hears appeals generated by District Courts within their jurisdiction. As a consequence of the geographical organization of courts, federal judges are required to maintain expertise in a wide variety of legal areas, and might be expected to learn different information on the ability of their colleagues in each case that they hear.

Most appellate cases are heard in panels of three judges, with a small number of cases heard by larger panels. While the process used to assign judges to cases is intended to be random in each circuit, the specific mechanism by which assignments are made varies by panel, and

includes methods such as the use of computer programs and the drawing of lots (Levy 2017). Some circuit courts, such as the fifth circuit choose panels in a manner that avoids having any judge serve too often with any other judge (Levy 2017), limiting the variance in exposure to other judges in our sample.

Appellate panels are composed of regular appellate judges, senior appellate judges (appellate judges who work part time and do not occupy congressional authorized positions), and visiting judges—typically either retired district, appellate, or supreme court judges or district judges serving on a district court that is subsidiary to the appellate court. Though the specific uses of visiting judges vary from court to court, circuits typically ensure that there are at least two regular appellate judges on each case, so visiting judges hear cases on panels with other visiting judges only in exceptional circumstances (Levy 2019). In addition, chief justices will often restrict the set of cases for which visiting judges are used, for instance by only allowing them to hear civil cases (Levy 2019). As a result, while the assignment of a specific judge to a specific case should not depend on that judge's characteristics for appellate judges or visiting judges, visiting judges as a group will be exposed to a different set of cases and colleagues than will regular appellate judges as a group. In addition, because visiting judges that hear cases throughout the year may not be available at the same time, the assumption that co-panelists are randomly assigned to visiting judges relies on the assumption that the timing of visits does not reflect judge characteristics. To avoid making this assumption, we restrict our sample only to appellate and senior appellate judges.

Once a panel is formed, judges accept legal briefs from the appellant and appellee, as well as occasional amicus briefs—briefs filed by people or organizations not directly involved in the

litigation but with an interest in the precedent set by the litigation. In some cases, panels also convene in person to hear oral arguments given by attorneys for appellants and appellees. After receiving briefs and hearing arguments, the panel confers, and decides to either affirm the decision of the district court, overrule the district court, send the case back to the district court to hear new evidence, or some combination of the above. A majority (typically two members) of a court must agree to this decision, and in 87% of cases, the court reaches a unanimous decision. Once a decision is made, a member of the court's majority is assigned to write the opinion of the court, and dissenting members of the panel write opinions explaining the grounds of their dissent.

Appellate court decisions create precedent that must be followed in lower courts. As a result, courts publish their opinions on cases in legal registers, making them available for reference and citation by other courts. However, because all litigants are entitled to appellate court review, many cases appear in court that break no legal ground and thus are not useful as precedent. As a result, courts only publish cases that, in the judgment of the court, include legal reasoning that is useful for citation. In 2017, fewer than 12% of Appellate cases were published. Our analysis examines the effect of the gender of co-panelists only on published cases, both because only published cases generate data and because these cases provide an opportunity to examine the legal reasoning and competence of colleagues in a way that routine cases do not.

2.2.1.1: Appointment of Judges and Clerks

Judges are appointed to both district and appellate courts by receiving a nomination from the President and confirmation from the Senate. Once appointed, judges have lifetime tenure on

the court, barring serious misconduct. Because judges hear cases in particular states, presidents tend to nominate judges recommended by the senators that represent the state in which a judge will hear cases, so long as that senator is from the president's party (Epstien et al, 2007). While there are no explicit legal requirements that federal judges hold specific qualifications, a majority of appellate court judges have prior judicial experience in either district courts or state courts, and 85% have prior experience practicing law (McMillan 2014). Likewise, while the process of nominating and confirming federal judges is often politically contentions, the large majority of nominees have been confirmed under all recent presidents, with George W. Bush holding the lowest confirmation rate at 78%, and Richard Nixon holding the highest Confirmation rate at 99% (Gramlich 2018).

Most Judges in our sample are over 60 years old, and have served in their current positions for more than a decade and a half. As a result, most federal judges were educated and began serving on the court when women were significantly less represented in the legal profession than they are today. While women make up half of new lawyers today, fewer than 5% of law school graduates were women prior to 1968, and only 36% were female in 1981, the year that the median federal judge in our sample finished law school (American Bar Association 2013). In 1992, when the average judge in our sample started their current position, women made up 43% of new law graduates. As a result, federal judges making hiring decisions now do so in a labor market in which women are significantly more numerous and successful than when the judges first formed impressions of the legal profession and of the judiciary.

Judges are provided with the budget for a staff, consisting of law clerks and administrative assistants. Typically hired directly out of law school, law clerks typically serve one-year or two-

year terms and are responsible for assisting judges with legal research and decision writing (Posner et. al. 2001). Appellate court clerkships are prestigious and highly sought-after positions, and are used as stepping stones to prestigious positions at top law firms, government agencies and the judiciary (Rhinehart 1994). Appellate court clerkships are especially valuable for law students hoping to enter the judiciary—every Supreme Court clerks serving from 2005-2014 had prior experience as Appellate court clerks (Hess 2015).

Despite efforts to push the hiring date for law clerks to the beginning of their third years, a substantial number of Law clerks are hired as early as the first semester of their second year of law school or are asked for informal agreements in the first semester of their second year, and begin work after their third year of law school (Posner 2007). As a result, there is as much as a two-year gap between the decision to hire a clerk and the clerk's start date. As a result of this, we assume that offers are made to clerks two years prior to their start date.

Due to the structure of the Appellate clerk market, Judges have wide latitude to choose the candidate clerk that best matches their preferences (Avery et. al. 2007). Because approximately half of all law third-year law students report applying to clerkship positions, appellate court clerks typically receive thousands of applications (NALP 2019). Offers of clerkships are typically made with very short decision windows (in some cases, as brief as 10 minutes), so most law students take the first clerkship offered.

Section 2.3: Data description

We pool data from several sources. Our two primary datasets are the "Judicial Yellow Books"—directories published by the *Leadership Library* that

list information on judge and clerks working in the US court system (Leadership Directories Inc, 2007-2014), and a directory of case records published by Federal Appellate Courts from 2007 to 2017 (Leagle, Inc, 2018). We also incorporate a rating of the conservatism of the president who appointed each judge using the DW Nominate algorithm (Epstein et. al. 2007).

Section 2.3.1: Primary Data

Judge and clerk information is collected from the "Judicial yellow books," published by the leadership library. Intended as a resource for attorneys presenting cases in State and Federal Court, the judicial yellow books include information on the names and backgrounds of judges serving in all levels of the federal court system, as well as the names and limited background information, of each judge's clerks. We purchased archived copies of the Judicial Yellow Books from the leadership library for the years 2007-2017, in the form of pdf pre-publication masters. We use these data to determine the characteristics of judges and the gender of the clerks hired by each judge in each year.

For each judge, the Yellow Books list current courthouse, start date, date of birth, appointing president, and provides information on all staff members.[11] It particular, for each staff member the judicial yellow books list name, title,

[11] The Yellow Books also contain information on the education (degrees earned, year of degree, and alma mater) and prior experience (in government, other judicial offices, law practice, private sector, military, and academia) of judges. In addition, they contain information on the education (degrees earned, year of degree, and alma mater) of judicial staff. This information, however, is largely incomplete, particularly for staff members, and is not used in our analysis. We use judge's college graduation year and law school graduation year to construct the age of the judge in case the date of birth is missing.

beginning and end of term, and contact information, and education. We use these data to determine which clerks worked for each judge in each year. We do so by taking advantage of the consistent formatting of the yellow book pages in the following way. Each subsection of the yellow books begins with a header "Chambers of Judge XXX," followed by a list of judge characteristics and a list of staff, preceded by a "Staff" sub-header. Law clerks and other staff are then listed with their title, followed by their name, with biographical information indented on following lines. We use this formatting both to determine which staff work for each judge and to exclude staff with titles other than "Law Clerk," such as "Administrative Assistant" or "Judicial Assistant." We count 7,443 court clerks, with between 1 and 4 clerks working for each judge in each year in 98% of cases.

The gender of judges and clerks are derived using the gender guesser algorithm (Pérez, 2016), a tool that determines if a name is male, female or uncertain by comparing it to a database of 40,000 names from 54 countries, primarily in the United States and Europe. We have verified the results of this algorithm both by comparing its results to the relative frequency of men and women with each first name in the US Census and by confirming the genders of judges with ambiguous names through web searches. Using this method, we are able to determine the gender of 95.5% of clerks and 91% of judges. We determined the gender of the remaining 9% of judges by examining biographical information from their court websites and/or news articles. Clerks for whom gender cannot be determined are excluded from the analysis.

Ethnicity of the judges is constructed by comparing judge names to surnames occurring at least 100 times in the 2010 decennial census. If more than 70% of census respondents with a judge's surname are Hispanic, we consider the judge Hispanic. If fewer than 30% of census respondents with a judge's surname are Hispanic, we consider the judge not Hispanic. We determined the ethnicity of judges with an indeterminate surname through court websites and news articles.

To determine which judges sat on panels together in each year, we scraped information from the online court records aggregator leagle.com. Leagle stores and categorizes the decisions handed published by the United States Courts - Trial, Appellate, and Supreme Courts. The library is comprehensive and contains over 5 million published and unpublished cases since 1950. We pool information on the universe of cases heard between 2007 and 2017, in total 50,813 cases. For each case, leagle provides the full text of the court's decision, exactly as it appears in published court documents. In addition, these court documents include a set of headers with the case's docket number and name, the date(s) the case was heard and decided, the court in which the case was heard, the names and affiliations of attorneys for the appellant and appellee, and most importantly for our purposes the names of the judges who heard that case. We use these records to identify the judges serving on appellate panels for a given case.[12] Specifically, we determine which judges heard a case by exploiting the

[12] Once a panel of judges is assigned to a case, the panel remains together until a decision is made. In rare cases, such as due in the case of sickness, a judge will be replaced on a panel. In these circumstances, all listed judges are included as interacting colleagues in our measures.

consistent formatting of Court document headers to identify the area where presiding judges are typically listed. We then capture each word in this section of the document to determine whether it is a possible judge name or a linking or descriptive word, such as “Before,” “Justice,” or “Honorable”. In following this procedure, we err on the side of including too many potential judge names rather than including too few. Therefore, we further compare the list of potential judge names to the list of surnames held by at least 100 people in the 2010 US Census, and remove all words not in the list of surnames.

We combine these two primary datasets by matching judges appearing in cases originating from each circuit court in each year in the case records data with judges listed in the judicial yellow book in that circuit court or a subsidiary district court in that year. Judges from subsidiary district courts are included because judges from District Courts are invited to serve on appellate panels (28 U.S. Code § 292). When merging the case data to the list of judges in the Yellow Books, 66% of potential judge names identified in the case records are also found in the list of judges in the relevant appellate court or subsidiary district courts. An analysis of the remaining 34% of potential names finds that they consist of names of attorneys or parties incorrectly categorized as judges or, in fewer cases, judges visiting from other circuits and retired judges hearing cases as senior judges. These 34% of names are dropped from the analysis and not used in determining the interactions of each judge. Among appellate court judges listed in the judicial yellow books, 85% of judges appear on at least one case record in the year that they are listed. Among the 15% of judges who do not appear in any

published cases, the majority are senior judges, and thus have discretion to hear few or no cases in a year. These judges may have heard cases over the year but heard no published cases, may have taken sick leave, or may be recorded inconsistently in the two data sources. Among district court judges listed in the judicial yellow books, 12% appear on at least one case record in the year that they are listed, consistent with a significant minority of district judges hearing appellate cases in any particular year. Judge names that do not match across these two data sources are eliminated. In total, we identify 298 Appellate Judges and 589 District Judges who served on an appellate court panel that produced a published opinion at least once between 2007-2017[13]. In our final database of 50,484 cases, 70% have three judges that are included in the analysis, 20% have two judges, 6% have one judge, and 4% have more than three judges. Cases can have fewer than three recognized judges if a member of the judicial panel is a visiting judge who is not from a subsidiary district court or if names are recorded improperly in our database. Cases can have more than three judges if they are heard en blanc (before all judges on an appellate court) or if a judge was replaced during the progress of the case due to illness or other circumstances.

Table C.1 compares the data on judges and cases used in this paper to official court statistics. First, we compare the count of published cases included

[13] Judges are identified in court documents by surname only. For nineteen surnames, multiple judges served simultaneously within a circuit (circuit court judges and district judges in subsidiary districts). For twelve of those surnames, the judges were of different genders. In these cases, interactions with a judge of these surnames was counted based on the "expected" gender of the judge. Because appellate judges hear, on average, 32 published cases per year, and district judges hear, on average, 0.5 published cases per year, we take the average gender of judges with each surname, assigning a weight of 32 to appellate judges and a weight of 0.5 to district judges.

in our data for each court with the count of published cases reported in annual Judicial Business Tables B.12 from November 2006 to November 2017. We recover between 89% and 98% of published cases for all circuit courts other than the third circuit. We recover only 69% of cases in the third circuit. Small deviations above and below the published numbers may be a consequence of January vs November date cutoffs, but it is likely that records from the Third circuit are incomplete. We also compare the number of judges appearing in the judicial yellow book and hearing cases in each year to the number of judges appearing in the federal judicial center database (Federal Judicial Center, 2019) serving in each year. There are an average of 268 Appellate Judges appearing in each year of our data, compared to 280 Appellate judges in each year of the federal judicial center data with start and end dates suggesting service in each year. The discrepancy between our data and the federal Judicial Center data is primarily a consequence of the fact that the judicial yellow books do not include all judges serving in appellate courts with senior status. In particular, in the fourth circuit, several judges are listed in the Judicial Center database as senior appellate judges who had never served as regular appellate judges—none of these judges appear in the judicial yellow book data. Likewise, judges who attained senior status prior to 1995 only occasionally appear in the judicial yellow book data. Because these judges also do not appear in our case records data, we believe that these judges have maintained senior status but are not actively hearing cases.

In addition to these core variables, we also analyzed the text of the court's decision reported in the Leagle library to extract further information on the cases heard. This information is not used in the primary analysis, but contributes to our analysis of heterogeneity of effect. First, we extracted information on how many times each case was cited by the supreme court, appellate courts, and district courts. In addition to the verbatim decisions of each court, Leagle collects and attaches a list of cases that cite each included case. We use this information to count the citations of each case, and to categorize them by court and year. Next, we determine the decision writer for each case using consistent formatting of decisions within appellate circuits. We also collect information on whether a dissent was filed in each case, whether oral arguments were conducted, and whether an amicus brief was filed, but do not use these variables in our analysis[14]. We use these citation rates to construct a measure of "quality" for each judge, defined as the average number of supreme court and appellate court citations for cases published by the judge, relative to the average rate of citation for cases published in the same circuit and year. While the citation rate of any particular case likely reflects the importance of that case as much or more as the quality of the decision writer, a high average citation rate is likely to indicate an ability to construct clear, convincing or novel legal arguments.

[14] We perform this textual analysis by taking advantage of the formatting of decisions and the presence of key words. We determine whether an amicus brief was cited in a case by searching the decision text for the term "amicus", and whether a concurring or dissenting was filed by searching for header text with the term "dissent, dissenting, dissents, concur, concurring, concurs" etc. We determine the decision writer using case formatting—opinions either begin with the decision writer's name or end with the writer's name. Citations are recorded in a standard bibliographic format, allowing us to both count citations and also determine which court and what level of court issued each citing opinion.

Section 2.3.2: Other Data

In addition to our two primary datasets, we incorporate two additional sources of information: the ratings of judge quality collected by the American Bar Association and an indicator of judge ideology, as measured by the DW-Nominate of the judge's appointing president.

The ratings of judge quality are used in our analysis as a secondary measure of judge quality. We use this for two purposes: (i) to investigate whether a judge qualification affects the responsiveness of a judge's hiring decisions to exposure to female colleagues, and (ii) to examine the effect of exposure to qualified female judges specifically (see Section 7). Each nominee to a position in the federal judiciary is rated by the American Bar Association Standing Committee of the Federal Judiciary on the basis of their professional qualifications. According to the rules of the Standing Committee, ratings are made on the basis of a judge's "integrity, professional competence and judicial temperament," and do not reflect the judge's "philosophy, political affiliation or ideology" (American Bar Association 2017). This Committee consists of 15 attorneys with standing to represent clients in appellate court circuits. Each member of the standing committee rates a nominee as either well-qualified[15], qualified, or not qualified, and the committee reports both the opinion of the majority and the opinion of the next largest minority (American Bar Association

[15] In the 1989-1990 term, judges could also be rated Exceptionally Well-Qualified. We collapse this category with the Well-Qualified category.

2017). We convert these ratings to numeric scores ranging from 0 (for a unanimous rating of not qualified) to 5 (for a unanimous rating of well-qualified). Because 68 judges were confirmed prior to the online dissemination of American Bar Association ratings, we include 482 judges with ratings of qualification.

We also include a measure of a judge's political ideology as both a control and as a source of heterogeneity in effect, taken from Epstein et. al. (2007). Presidents typically defer to senators from their own party on the nomination of a judge from a senator's state. Epstein et. al. exploit this senatorial courtesy to assign a judge the ideology of same-party home-state senators, when such exist, and the ideology of nominating presidents when home-state senators are of a different party than the president. The ideology of presidents and senators are based on the record of votes (for senators) and stated support (for presidents), calculated using the DW-Nominate algorithm (Lewis et. al. 2017)[16].

Section 2.3.3: Sample Selection

We identify 365 distinct appellate judges in the judicial yearbook data, of whom 215 hear at least one appellate case in at least one year of the data (the rest consist of inactive senior judges). If we had records for all eight years for each of these 298 judges, we would have a total sample of 2384 observations. In reality,

[16] DW-Nominate scores are computed by examining the likelihood that each member of congress votes with each other member of congress. The DW-Nominate algorithm assumes that the likelihood that each voter votes in favor of a piece of legislation is determined by the distance of that bill's ideological content from the voter's ideal point on a three-dimensional ideology space. The algorithm determines a combination of vote ideological content and voter ideology using maximum likelihood estimation. The algorithm treats stated support or opposition to votes by presidents as votes.

however, 20% of judges start after 2007, and another 15% retire before 2014. Of those who started prior to 2007 and continued in their positions until 2014, 5% heard no cases during one year of their service. As a result, only 63% of judges in our sample appear on appellate panels in each year, and we only observe 2158 judge-years of interaction on federal appellate court. Furthermore, 83 judges who hear cases in at least one year do not hire any appellate clerks during the sample period, and all but 8% of the remaining 215 judges have at least one year where they hire no clerks. As a result, our final sample includes 1074 observations, at the judge by year level, from 215 judges over eight years. As shown in Table C.2, our sample consists of judges in years where the judge was on at least one panel with a published case, hired at least one clerk in the following year, and is not missing any primary covariates. We also include regressions that control for the current gender composition of a judge's law clerks—this covariate is missing when a judge has no law clerks on staff, resulting in missing values for 87 observations, primarily in the first year of a judge's tenure. Characteristics of this sample are available in appendix Table C.3. As shown in Table C.3, 26% of observations come from female judges. Judges hire an average of 2.9 law clerks per year, and hire at least one female law clerk in 70% of years in which they make a hiring decision. Overall, 42% of clerks are female.

Section 2.4: Empirical model and identification strategy

Our empirical strategy takes advantage of the random assignment of judges to panels to regress a measure of interaction with female judges in a given year on the likelihood of hiring at least one female clerk in the following year. Because assignment of judges to cases is random conditional on circuit, year, and the status of the judge (regular appellate judge or visiting district judge), we control for fixed effects at the court by year by district judge level. Variation in the fraction of co-panelists who are female in a year, conditional on court, year, and judge type, is due entirely to the random assignment of judges to cases, and to the determination of panels that a case is worthy of publication. We estimate the effect of exposure to female colleagues on the likelihood of hiring a female clerk using the following regression equation:

$$Hire_{j,c,t+1} = \beta Inf_{j,c,t} + \delta X_{j,c,t} + \theta_{c,t} + \varepsilon_{j,c,t} \quad (1)$$

Where $Hire_{j,c,t+1}$ is an indicator of whether judge *j*, in court *c*, hired at least one female clerk at time *t+1*; $inf_{j,c,t}$ is exposure to female judges, $X_{j,c,t}$ is a set of judge characteristics, and $\theta_{c,t}$ is a set of court by year fixed effects.

We measure exposure to female judges $inf_{j,c,t}$ as the fraction of co-panelists on the cases heard by judge *j* in year *t* that are female. There are a few noteworthy characteristics of this measure. First, because we calculate the fraction of co-panelists who are female, our measure of exposure to female colleagues does not depend on the volume of cases heard by a judge in a particular year. This decision is justified both by the assumption that full-time judges with few cases are likely to have more time-consuming cases than those with many cases, and by the

assumption that the salience if individual cases is likely greater when a judge has heard fewer cases in a year. Second, this measure does not distinguish between interactions with a single female colleague on many cases and interactions with multiple female colleagues, each on an individual case. This decision is justified by a learning model in which each case heard with a co-panelist reveals a small amount of information about that co-panelist's ability, which is then used to inform the likely distribution of legal professionals with the co-panelist's gender. If there is substantial uncertainty about the ability of each judge, repeated interaction with one judge will have similar information content to individual interactions with multiple judges.

We measure propensity to hire female clerks via an indicator of whether any female clerk was hired in the year following a judge's exposure. This gives equal weight to judges who hire one clerk in a year (as do most) or two clerks in a year.

In order to demonstrate the feasibility of this approach, we estimate the variation in our main dependent and independent variables that is not accounted for by court by year variation. As shown in Table C.4, very little variation in either the hiring decisions of judges or in the exposure of judges to female colleagues is explained by differences between courts and years. Likewise, observed judge characteristics do not explain a significant amount of variation in either judge hiring decisions or exposure to female judge.

Section 2.4.1: Test of Identification

Our key identifying assumption in this paper is that variations in the gender composition of co-panelists within a particular circuit and year is unrelated to a judge's preferences for female clerks and a judge's available labor pool. This assumption is justified by the assertion, common

to all appellate circuits, that judges are randomly assigned to cases (Abramowicz and Stearns 2005). While recent literature has cast doubt on the strict random assignment of cases to judges, violations to random assignment are small, unlikely to be sustained over a year, seen only in a few courts, and unlikely to be related to judge's preferences or labor pools. In particular, work by Chilton and Levy (2015) found that due to scheduling conflicts and similar concerns, the assignment of judges to appellate panels deviates from random assignment in several courts. As a consequence, the distribution of Republican appointees across court differs from what would be expected by chance slightly in the second, sixth and DC circuits, and more substantially in the ninth circuit. However, the likelihood that a republican will serve with another republican differs from chance by less than a percentage point in all circuits but the second and ninth, in which it differs from chance by less than two percentage points. In addition, as shown in Table 2.1, judges appointed by Republican presidents are as likely to serve on panels with female judges as are judges appointed by Democratic presidents, indicating that any non-randomness in the assignment of judges to panels on the basis of political party does not affect the likelihood of serving on panels with female colleagues. Finally, as shown in Table 2.2, the inclusion of controls for judge characteristics, including party, does not weaken the measured effect of exposure to female colleagues on hiring decisions. Levy (2017) examines a broader range of potentially non-random scheduling decisions made by the chief judge's office of each appellate circuit, finding, for instance, that one circuit had a tradition of ensuring that judges have the opportunity to be the presiding judge on one case in their first year by constructing a panel with two senior or visiting judges. However, these deviations from strict randomness are small enough that federal judges themselves believe panels to be randomly constructed (Levy 2017).

We test the potential threat of nonrandom case assignment to identification across a number of dimensions by regressing our main independent variable, the fraction of a judge's co-panelists who are female in each year, onto a series of observed judge characteristics—specifically, on a judge's Hispanic ethnicity, quadratic of years of experience, quadratic of age, political party, quadratic of ideology of nominating president, and gender composition of current staff, controlling for judge gender and for court by year fixed effects. As shown in Table 2.1, there is little to no relationship between the exposure of a judge to female colleagues and any observed judge characteristics.

Table 2.1: Balance Tests for Random Assignment to Panels

Dependent Variable: % co-panelists who are female	(1)	(2)	(3)	(4)
Age	0.0043 (0.0053)	0.0016 (0.0035)	0.0012 (0.0041)	0.0051 (0.0065)
Years on Current Court	0.0104** (0.0052)	-0.0008 (0.0032)	-0.0003 (0.0038)	-0.0004 (0.0064)
Ideology Score	0.0343** (0.0169)	-0.0008 (0.0106)	-0.0043 (0.0126)	0.0188 (0.0254)
Republican	0.017 (0.0114)	0.0011 (0.0068)	-0.0014 (0.0079)	0.0118 (0.017)
% of current staff female	-0.0101 (0.0152)	0.0098 (0.011)	0.0185 (0.0137)	-0.0137 (0.0237)
Hispanic	0.0063 (0.0252)	-0.0148 (0.0124)	-0.0126 (0.0116)	-0.0748** (0.0341)
Court X Year fixed effects	No	Yes	Yes	Yes
Sample	All	All	Male	Female
Observations	1074	1074	795	279

Notes: Table reports coefficients from a series of regressions of the fraction of co-panelists who were female in a year on a series of judge characteristics. Column (2) controls for whether the judge is female. Columns (2)-(4) include court by year fixed effects. Column (3) shows regressions on male judges, column (4) shows regressions on female judges. Significance levels are: * 10%, ** 5%, *** 1%.
Source: Case dataset collected by authors (see data section for details).

Table 2.1 shows the relationship between a variety of judge characteristics and the main variable of interest, for both the full sample (columns 1 and 2) and separately for male and female judges (columns 3 and 4). We separate the sample by judge gender because the expected gender composition of colleagues is mechanically lower for female than for male judges, due to the fact that judges cannot interact with themselves. While we control for judge gender in column (2), the size of this mechanical effect is larger in small circuits such as the first circuit (with 10 judges) than in the ninth circuit (with 48 judges). Overall, the relationships we observe between judge characteristics and interaction with female colleagues is no greater than would be expected by chance, with the only statistically significant relationship being a lower likelihood of serving with female colleagues for female Hispanic judges at the 5% level. Because we perform 18 tests, a single test that is significant at the 5% level would be expected even if there were no true relationship between any of the judge characteristics and interactions with female colleagues.

Section 2.5: Main Results

Table 2.2 presents ordinary least squares (OLS) regressions in which the dependent variable is an indicator of whether a judge hired a female clerk in year *t+1* and the key independent variable is the fraction of the judge's co-panelists who were female in year *t*. Column (1) includes court by year by district judge fixed effects, but no additional covariates. Column (2) adds controls for judge gender, hispanic ethnicity, and age, and column (3) adds controls for the political party of the judge's nominating president, a quadratic of the DW-Nominate score of the judge's nominating president, and a quadratic of the judge's years of experience on their current court. This table shows that a one standard-deviation increase in a judge's exposure to female

colleagues, or an increase of 0.11 in the fraction of judicial interactions with female colleagues, leads to a 4 percentage-point increase in the likelihood that a judge hires a female clerk. The addition of additional controls very slightly increases the precision of the estimate but has no detectable effect on the magnitude of the estimate.

Table 2.2: Effect of Serving with Female Judges on Hiring Decisions

Dep Var: Probability of hiring any female clerk in next year	(1)	(2)	(3)	(4)
Fraction of co-panelists who are female	0.3892**	0.4296**	0.3972**	0.4210**
	(0.1918)	(0.1867)	(0.1845)	(0.2036)
Female	0.0782**	0.0658	0.0352	0.0550
	(0.0395)	(0.0406)	(0.0384)	(0.0395)
Hispanic		0.2002***	0.1674***	0.1582***
		(0.0368)	(0.0464)	(0.0481)
Age		0.0141	0.0315	0.1156
		(0.1642)	(0.1817)	(0.2186)
Age^2		-0.0001	-0.0000	-0.0001
		(0.0001)	(0.0001)	(0.0002)
Years on Current Court			-0.0749	-0.0696
			(0.0617)	(0.0671)
Years on Current Court^2			0.0101	-0.0069
			(0.1227)	(0.1292)
Ideology Score			-0.1774	-0.1385
			(0.1856)	(0.1943)
Ideology Score^2			0.0437	0.0113
			(0.2453)	(0.2488)
Republican			0.0101	-0.0069
			(0.1227)	(0.1292)
% of current staff female				0.1721***
				(0.0527)
Other co-panelist characteristics	No	No	No	No
Court X Year X District Judge FE	Yes	Yes	Yes	Yes
Observations	1074	1074	1074	987
Dependent Variable Mean	0.7030	0.7030	0.7030	0.6961

Notes: This table reports estimated coefficients from the regressions described in equations (1) and (2) in the text. The dependent variable is an indicator of whether a judge hired at least one female clerk in the following year, conditional on hiring any clerk. The table reports the regression of the dependent variable on the fraction of co-panelists who were female in each year. Column (5) adds controls for the fraction of co-panelists who are republican, who are younger than 60 years old, who have served fewer then 10 years on the court, who have a current staff that is more than 50% female, and who have an above average citation rate. Significance levels are: * 10%, ** 5%, *** 1%. Source: Judicial yellow books, case dataset collected by authors (see data section).

Section 2.6: Placebo Regressions and Robustness

We explore a few additional sources of potential concern with our identification strategy. First, one might be concerned that a judge's propensity to hire female clerks changes over time in a manner that is correlated to their exposure to female judges. For example, if judges with greater seniority have more flexibility in scheduling vacation days, a judge might increasingly take vacations that are scheduled while (predominantly non-senior) female judges are working and have a higher chance of hiring their (predominantly male) favored clerk candidates. To test this hypothesis, we regress a judge's new staff in one year on their exposure to female colleagues in the same year. Because staff are hired one to two years in advance of their start-date, this exposure cannot have a causal effect on hiring. As shown in columns 1, 2, and 3 of Table 2.3, exposure to female clerks in the year following a hiring decision is unrelated to that hiring decision[17].

Second, one might worry that a judge's ideology and/or experience affects the labor pool from which they hire, either because judges prefer ideologically similar clerks or because more senior or more ideologically mainstream judges can offer more prestigious positions and are thus able to hire the most sought-after clerks. If the status of a particular group of judges (say, conservative judges) within a court covaries with the number of female clerks available within the appropriate local labor market (say, members of conservative judicial organizations), this could lead to spurious correlation between the hiring of female clerks and the number of female

[17] Note that the sample size for these comparisons of time *t* exposure to female colleagues on time *t-1* staffing starts have a higher sample size than do comparisons of time *t* exposure to female colleagues on time *t+1* hiring decisions. Because hiring decisions happen two years prior to staffing starts, staff hired after 2014 cases are heard appear in our data in 2017, requiring us to drop case data from 2015-2017. In contrast, the placebo test only requires us to drop case data from 2007.

co-panelists. We address this concern by matching each judge to the most similar judge within their court (with replacement) and regressing the exposure and characteristics of each judge to the hiring decisions of their match. We determine matches by regressing the fraction of each judge's staff who are female onto the judge's characteristics and court. We then select the judge with the most similar predicted staff gender composition. As shown in columns 4, 5 and 6 of Table 2.3, while there is a positive association between the exposure of a judge's most similar colleague and their likelihood of hiring a woman, the relationship is not statistically significant.

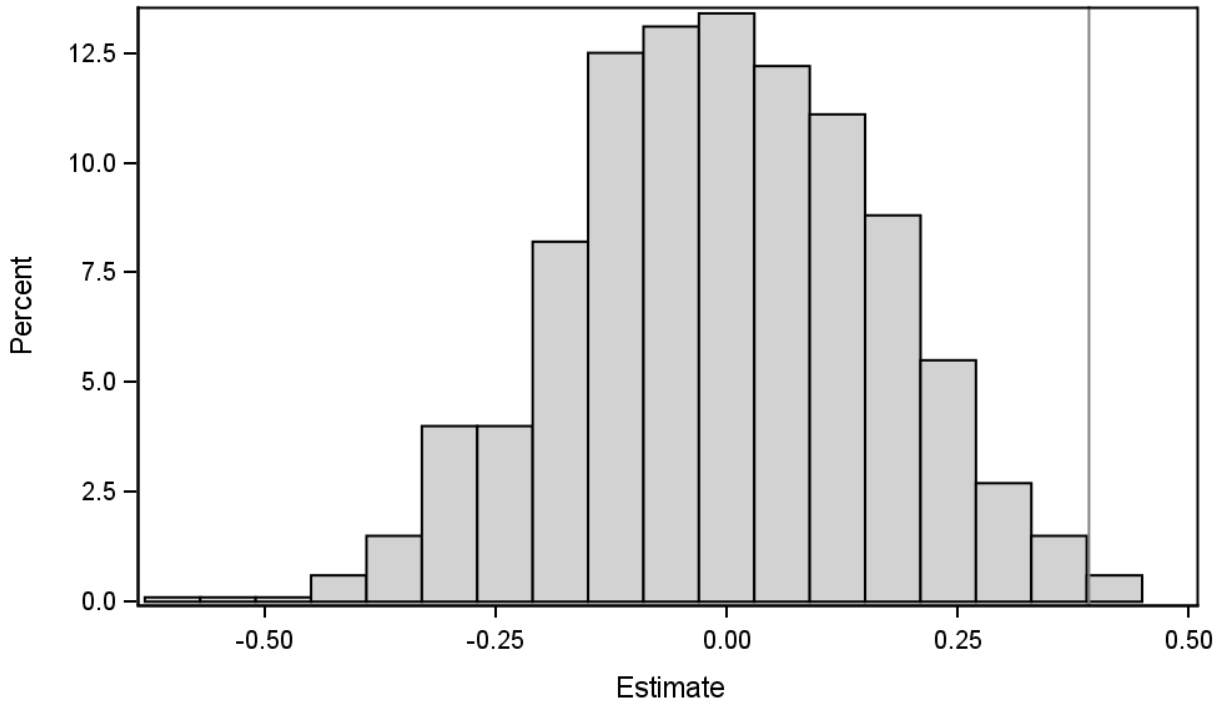
Table 2.3: Placebo Tests

	Prob. **Actual Judge** Hired any female clerks in **past** year			Prob. **Matched judge** hired any female clerks in **next year**		
	(1)	(2)	(3)	(4)	(5)	(6)
fraction of co-panelists who are female	0.0070	0.1120	-0.0139	0.1777	0.1832	0.1353
	(0.1581)	(0.1613)	(0.1135)	(0.1810)	(0.1791)	(0.1818)
Female		0.1700***	0.1095**		0.1823***	0.1538***
		(0.0356)	(0.0472)		(0.0368)	(0.0495)
Hispanic		0.1105	-0.0146		0.0162	0.0982
		(0.1357)	(0.1158)		(0.1953)	(0.2167)
Age			-0.0533			-0.1499
			(0.0933)			(0.1916)
Age^2			0.0055			0.0065
			(0.0093)			(0.0181)
Years on Current Court			-0.0837**			-0.0726
			(0.0407)			(0.0671)
Ideology Score		-0.0001	0.0000		-0.0001	-0.0001
		(0.0001)	(0.0001)		(0.0002)	(0.0002)
Ideology Score^2			-0.1668			0.0322
			(0.1698)			(0.2513)
Republican			0.0189			0.0019
			(0.0574)			(0.1274)
Years on Current Court^2			0.9783***			0.1752***
			(0.0334)			(0.0524)
% of current staff female	X	X	X	X	X	X
Court by Year FE	Yes	Yes	Yes	Yes	Yes	yes
Observations	1427	1427	1413	1046	1046	986
Dependent Variable Mean	0.70	0.70	0.70	0.70	0.70	0.70

Notes: This table reports OLS estimated coefficients from the regression described in equation 3 are reported from two placebo tests. Standard errors are robust and clustered at the judge level. In columns 1-3, the dependent variable is an indicator of whether the judge hired at least one female clerk in year *t-1.* In columns 4-6, the dependent variable is an indicator of whether the most similar judge within a court, based on characteristics predicting the employment of female clerks, hired at least one female clerk in year *t+1*. The table reports the regression of the dependent variable on the fraction of co-panelists who are female in each year. Judge characteristics include quadratics of judge age, experience in current position and ideology, judge gender, hispanic ethnicity and party of nominating president. Significance levels are: * 10%, ** 5%, *** 1%. Source: Judicial yellow books, case dataset collected by authors (see data section).

Finally, we perform a permutation resampling procedure to determine whether our standard errors accurately reflect the distribution of likely effect sizes. To do this, we randomly reassign hiring decisions to judges within each court and year, and estimate the full model (shown in Table 2.2, column 3) for each random assignment. Figure 2.1 shows the distribution of 2000 randomly generated effects against the estimated effect, and shows that our results are unlikely to have occurred by chance.

Figure 2.1: Randomization-Based Inference for Fraction of Co-Panelists who are Female



Note: Figure shows distribution of coefficients obtained from the final OLS specification in Table 3 while replacing the fraction of co-panelists who are female for a judge with the fraction of co-panelists who are female from a randomly selected judge from the same court and year. Vertical line represents the actual estimate obtained in the final specification in Table 3.

Section 2.7: Heterogeneity

We next examine heterogeneity in the effect of interaction with female colleagues by both the characteristics of the influenced judge and the characteristics of the interacting female colleagues. We examine seven sets of characteristics, examining both whether judges with each characteristic are more affected by interactions with female colleagues and whether female colleagues with each characteristic have a greater effect when serving as co-panelists. These characteristics are: judge gender, judge quality, as measured by rate at which a judge's decisions are cited, relative to other decisions from the same court and year, the fraction of a judge's staff currently composed of females, judge age, judge experience, the political party of the judge's nominating president, and whether the judge is visiting from a district court. As shown in Table 2.4, we find suggestive evidence that male judges, judges who's current clerks are less than 50% female, and judges with fewer average citations are more influenced by interaction with female colleagues than are female judges, judges who's current clerks are more than 50% female, and highly cited judges. These findings are consistent with a model of judge learning, where judges who are most likely to be surprised by competent female colleagues showing the largest changes in hiring.

Table 2.4: Heterogeneity in Main Effect

Dep Var: Probability of hiring any female clerk in next year

Panel A: Characteristics of Judge

Var. Z:	Female (1)	>50% Fem Staff (2)	Abv-Avg Citations (3)	Republican (4)	Age < 60 (5)	< 10 yrs on court (6)
frac co-panelists female	0.4309**	0.4631**	0.4131**	0.2665	0.3415*	0.2950
	(0.2002)	(0.2345)	(0.1958)	(0.2303)	(0.2004)	(0.2093)
frac co-panelists female X Variable Z	-0.2720	-0.2038	-0.4161	0.1402	0.0186	0.1747
	(0.2858)	(0.2759)	(0.3445)	(0.2842)	(0.2534)	(0.2599)
Variable Z	0.0838	0.1247*	0.1561*	-0.0478	0.0677	-0.0202
	(0.0719)	(0.0700)	(0.0870)	(0.1514)	(0.0807)	(0.0862)
Court by Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1074	1074	1074	1074	1074	1074
Dependent Variable Mean	0.7030	0.7030	0.7030	0.7030	0.7030	0.7030

Panel B: Characteristics of co-panelists

Var. Z:	>50% Fem Staff (2)	Abv-Avg Citations (3)	Republican (4)	Age < 60 (5)	< 10 yrs on court (6)	District Judge (7)
fracco-panelists female	0.4655**	0.3255	0.5241***	0.2872	0.3788*	0.3376*
	(0.2152)	(0.2248)	(0.2013)	(0.2854)	(0.2177)	(0.1839)
frac co-panelists female AND variable z = yes	-0.2280	0.0799	-0.4536	0.0985	-0.0790	1.8873
	(0.2441)	(0.3476)	(0.3127)	(0.3030)	(0.2956)	(2.4469)
Court by Year FE	Yes	Yes	Yes	Yes	Yes	No
Observations	1074	987	1074	1074	1074	1074
Dependent Variable Mean	0.70298	0.69605	0.70298	0.70298	0.70298	0.70298

Notes: Panel A reports estimated coefficients on the interaction of the main independent variable with 7 interaction variables describing judge characteristics: Column (1) is an indicator of whether the judge is female. Column (2) is the fraction of the judge's staff that is female in year t, prior to the new hire. Column (3) is an indicator for whether the judge's decisions are cited more often than expected based on court and year. Column (4) is an indicator of whether the judge was appointed by a republican president. Column (5) is the age of the judge (decades). Column (6) is the decades of experience of the judge. Column (7) is whether the judge is a district judge. Panel B reports the estimated coefficients on the interaction of the main independent variable with 6 characteristics of interacting judges. The main independent variable in each regression is the fraction of a judge's co-panelists who are female in each year. All results control for quadratics of judge age, experience in current position and ideology, judge gender, hispanic ethnicity, party of nominating president, and the current fraction of the judge's staff who are female. Significance levels are: * 10%, ** 5%, *** 1%. Source: Judicial yellow books, case dataset collected by authors (see data section).

Panel A of Table 2.4 shows the relationship between a judge's characteristics and the effect of interaction with female colleagues on hiring. As shown in column 1 of Table 2.4, an 10 percentage-point increase in exposure to female colleagues is expected to increase the likelihood that a male judge hires at least one female clerk by 4.3 percentage points, but to increase the likelihood that a female judge hires at least one female clerk by only 1.6 percentage points, though this difference in effect may be the result of chance. This suggestive finding is consistent with the hypothesis that female judges may hold more positive prior beliefs about the availability of talented female clerks than do male judges, and thus may be change their beliefs less in the presence of female colleagues.

Likewise, column 2 of Panel A suggests that judges who's current clerks are 50% female or more are less affected by interactions with female judges than are judges who's current clerks are less than 50% female. A 10% increase in the female fraction of copanelists is expected to increase the likelihood of hiring a female clerk by 4.6 percentage points for judges with majority-male clerks, compared to 2.6 percentage points for judges with at least 50% female clerks. Together, these findings suggest that the judge's most affected by interaction with female colleagues are those most likely to be surprised to interact with capable female colleagues.

Column 3 of panel A suggests that judge's who's decisions are cited at an above-average rate are less affected by interactions with female colleagues than are judges who's decisions are cited at a below average rate. A 10 percentage-point increase in the female fraction of co-panelists increases the likelihood that a less-cited judge hires a female clerk by 4 percentage points, and that a more-cited judge hires a female clerk by 0 percentage points. While this difference is no larger than would be expected by chance, it may reflect differences in the labor

pools available to more versus less highly regarded judges, with highly cited judges able to hire their first choice candidates.

Panel B shows the relationship between the characteristics of female co-panelists and the effect of serving with female co-panelists on hiring. While none of these interactions are statistically significant, we find suggestive evidence that interactions with female district court judges and with female Democrats have larger effects on hiring decisions than do interactions with female Republicans or female Appellate judges. In particular, a 10 percentage-point increase in the fraction of co-panelists who are female democrats is expected to increase a judge's likelihood of hiring a female clerk by 5.2 percentage points, while a 10 percentage-point increase in the fraction of co-panelists who are female republicans will only increase a judge's likelihood of hiring a female clerk by 0.7 percentage points. The greater influence of Democratic female judges is somewhat surprising, since female republican judges are rarer and thus perhaps more surprising and informative. However, the differences are small enough that they can be attributed to chance. Meanwhile, a 10 percentage-point increase in the fraction of co-panelists who are female district judges is expected to increase the likelihood of hiring a female clerk by 22.2 percentage points, while a 10 percentage-point increase in the fraction of co-panelists who are female appellate judges is only expected to increase the likelihood of hiring a female clerk by 3.4 percentage points. The particular impact of interactions with female district judges may reflect the fact that district judge's are less well known to appellate judges than are their appellate colleagues, so a given interaction with a district judge is likely to carry more information than would an interaction with an appellate judge.

Finally, column (2) of panel B shows no evidence that interacting with female judges with female staff has a greater effect on hiring than does interacting with female judges with fewer female staff. This makes it somewhat less likely that we are capturing an effect of referrals and information about specific candidates conveyed from female judges, rather than changes in judge's perceptions of women in law.

Section 2.8: Conclusion

This paper presents evidence that federal judges are more likely to hire female clerks after serving on a panel with female judges. In particular, we find that a one standard deviation increase in the fraction of published cases heard alongside female colleagues increases a judge's likelihood of hiring at least one female clerk by 3.5 to 4 percentage points. Because judges are broadly unconstrained in who they hire as a clerk, we interpret this change in hiring practices as a change in judge's assessment of the ability of women in judicial practice and law.

This finding suggests that increases in the diversity of the upper rungs of a profession can shift attitudes in a way that creates opportunities at the entry level of a profession. This in turn suggests that policies aimed at increasing the diversity in the leadership of a profession, such as affirmative action policies or policies requiring that a certain number of board seats be filled by women, may have benefits beyond their immediate beneficiaries.

Because this work is unique in its ability to estimate the role of peer interactions among established professionals, it is reasonable to wonder whether particular characteristics of the legal profession or the judiciary make it more or less susceptible to such effects. One possibility is that the effect of peer exposure may be particularly salient in occupations which have seen

dramatic increases in the number of women at the entry level and in higher positions over the career of the current leadership generation. Law is such a profession—while fewer than 5% of American law school graduates were women prior to 1970, 50% of recent law school cohorts have been female (American Bar Association 2013). As a result of this generational change in law, the great majority of federal judges began their careers in a cohort composed predominantly of men, and at a time when few or no women occupied prominent positions in law, such as professorships or judicial appointments. It is thus reasonable to suspect that these judges have developed expectations about women in the profession that do not reflect the characteristics of the newest cohort of lawyers. Our results may be large in law precisely because judges have developed attitudes that can be counteracted with new evidence.

We may also expect the legal profession to exhibit large effects of peer interaction because of a lack of clear and objective criteria for hiring decisions. As argued in Goldin (2015), negative perceptions of women's abilities within an occupation may be less persistent in occupations where clear tests of abilities are available. While certain credentials are very helpful in securing employment in law, such as graduation from a top law school or membership on a legal review, the legal profession lacks highly objective measures of quality or productivity available in some engineering fields.

CHAPTER 3: The Effect of Children's Safety on Parenting Strategy

ABSTRACT

I develop and test a model that explains differences in parenting style by socioeconomic status. Spanking, severe discipline, and other forms of "Authoritarian Parenting" are more common among low income and black parents than among high income and white and Hispanic parents. Although they are associated with increased short-term obedience, these "Authoritarian" parenting strategies are also associated with lower levels of cognitive development, self-esteem, school performance and "moral internalization." I construct a multi-stage parenting model in which children choose levels of school effort and delinquent behavior while heavily discounting future consequences, and parents altruistically regulate their children both by disciplining them and investing in their self-control. The model predicts that parents employ more discipline when the negative effects of child delinquent behavior are large. I test this model by measuring the effect of school safety (which influences the cost of child delinquent behavior) on parenting practices, measured in the Los Angeles Family and Neighborhood Survey (LA.FANS). Because school safety affects parents only through its effect on children, the relationship between school safety and parenting style should not reflect parents' stress, social isolation, or economic circumstances. Controlling for family, neighborhood and school characteristics, I find that a 1 standard deviation increase in school disorder is associated with a 0.11 standard deviation increase in harsh parental discipline, with larger effects for poor and Black households. I argue that this effect is driven by parents' concern about the cost of their child's misbehavior.

JEL Codes: J13,C70

Keywords: Child Care, Children, Youth, Parenting Style, Human Capital, Altruism, Intergenerational Preference Transmission

Section 3.I: Introduction

> *My father was so very afraid. I felt it in the sting of his black leather belt, which he applied with more anxiety than anger, my father who beat me as if someone might steal me away, because that is exactly what was happening all around us. Everyone had lost a child, somehow, to the streets, to jail, to drugs, to guns. It was said that these lost girls were sweet as honey and would not hurt a fly. It was said that these lost boys had just received a GED and had begun to turn their lives around. And now they were gone, and their legacy was a great fear.*

-Ta-Nahesi Coates; *Between the World and Me*

> *You forget the reason I ride you so hard and give you so much shit is because I love you. Everything I have ever done I've done from a place of love. If I don't punish you, the world will punish you even worse. The world doesn't love you. If the police get you, the police don't love you. When I beat you, I'm trying to save you. When they beat you, they're trying to kill you.*

-Trevor Noah's mom; *Born a Crime*

Poor parents in the United States interact with their children very differently than do affluent parents. Low-income parents are more likely to report using physical punishment on their children than are high income parents (Pinderhughes et al, 2000) and are more likely to describe an ideal child as "obedient" (Park and Lau 2016). Within income groups, parents who raise children in more challenging environments also report more use of physical punishment, with Zhang and Anderson (2009) finding that parents exposed to moderate to high levels of neighborhood violence are more than twice as likely to use physically aggressive parenting approaches than are similar parents exposed to low levels of neighborhood violence. Similar differences exist by race, with Black and Hispanic parents more likely than White parents to demand respect

for parental authority (Dixon, Graber, and Brooks-Gunn 2011), and Black parents using more discipline on average than white or Hispanic parents, both physical (Berlin 2009) and non-physical (Weinberg 2001).

A body of theory and evidence in developmental psychology, first developed by Baumrind (1967), connects these practices into a set of parenting "styles"—combinations of goals and practices that characterize the parent-child relationship. While "authoritarian" parents value respect and obedience, expect their instructions to be obeyed because they are in charge, and use punishment to keep their children in line, "authoritative" parents value critical thinking and development, and attempt to have their children negotiate decisions within boundaries and safeguards developed jointly between parent and child. While authoritative parenting has become more common among all races and socioeconomic statuses since Baumrind's typology was developed (Straus and Donelley 1994), poor parents, black parents, and less-educated parents have persistently employed more authoritarian parenting strategies than have wealthy parents, white parents, and highly-educated parents (Berlin et. al. 2009).

The effects of these parenting differences may extend well past childhood. Studies examining long-run differences between children raised in an authoritarian style and children raised in an authoritative style have found that harsh discipline is associated with lower school performance (Dornbusch et al 1987; Steinberg et al 1992), less "moral internalization" (Martinez 2007; Gershoff and Grogan-Kaylor 2016), lower self-esteem (Lamborn et. al. 2008) and lower earnings (Dornbusch et al 1987). This evidence of the negative effects of harsh

discipline and authoritarian parenting have led the American Society for Pediatrics to recommend against any forms of physical punishment for children (Siege and Siegel 2019), and have spurred multiple efforts to find policies that reduce the use of harsh discipline among poor families, including the "baby college" program for young parents run by the Harlem Children's Zone (Harlem Children's Zone, 2014) and Nurse Family Home Visiting Programs (Howard and Brooks-Gunn 2009).

These findings have raised the question of why some parents would practice forms of parenting with significant negative long-run consequences for children, and several economists have developed models to explain harsh or authoritarian parenting among low-income households. Weinberg (2001) and Cobb-Clark, Salamanca and Zhu (2016) argue that authoritarian parenting requires fewer resources than does authoritative parenting, and thus is preferred by disadvantaged parents. Weinberg (2001) argues that low-income parents use harsh and physical punishments at high rates because they cannot withhold allowances. Cobb-Clark, Salamanca and Zhu (2016) hypothesizes that authoritative parenting strategies require greater cognitive effort than do authoritarian parenting styles, and that low-income parents are less likely to use such parenting styles because their mental reserves are depleted by other sources of stress. Similarly, Cunha (2015) argues that poor parents are less aware of the long-term costs of harsh discipline than are rich parents, and are thus less likely to take on the higher cognitive burden demanded by authoritative parenting styles. Finally, Doepke and Zilibotti (2017) claim that the benefit of

authoritative parenting comes largely from the ability of children to make their own decisions with guidance from their parents, and that this benefit is larger for children with a broader range of adult opportunities.

In this paper, I propose an alternative model. Even if low-income parents were fully aware of the long-run costs of harsh discipline, had the mental and emotional reserves to choose the parenting style they thought best for their kids, and expected that authoritative parenting would help their children in the long run, they still may use more harsh discipline with their kids when they expect the costs of bad decisions during childhood to be especially high. Because physical punishment is effective in getting children to immediately comply with instructions (Gershoff and Grogan-Kaylor 2016), it is an effective tool for keeping kids out of trouble in the short-run. And if short-run trouble includes serious consequences, like physical harm or arrest, short-run compliance might outweigh other considerations.

To test this hypothesis, I examine the effect of a change in the safety of a child's school on parents' disciplinary practices. I use the Los Angeles Family and Neighborhood Survey (Pebley and Sastry 2011), a panel survey of 2500 Los Angeles County families conducted in 2000 and 2006, to measure the effect of a change in school safety on parenting decisions. This dataset is unique in including precise geographical information on sampled households alongside detailed information on parent-child interactions, from the perspectives of the parent and the child. I regress several measures of parental discipline onto the safety of a child's school, controlling for family, neighborhood, and school

characteristics. I find evidence that a 1 standard deviation increase in a hazardous school index increases parent's intention to use harsh discipline by 0.1 standard deviations. For poor and Black households, hazardous schools also result in more actual discipline, with an effect size of 0.11 for poor households and 0.17 for Black households. This effect size suggests the difference in the use of discipline between a black household attending an average school and a black household attending a school one standard deviation above the mean in disorder is approximately half of the difference in discipline between black and white parents. Likewise, the difference in the use of discipline between a poor household attending an average school and a poor household attending a school that is one standard deviation above the mean in disorder is more than 2/3 the difference in discipline between poor and non-poor parents.

This paper makes two contributions to the literature. First, I propose the first economic model of parenting style that centers on the conditions of childhood outside of the family, rather than on the limited resources or knowledge of parents (Weinberg 2001; Cobb-Clark Salamanca and Zhu 2016), or on the economic conditions and opportunities of adults (Doepke and Zilibotti 2017). This model suggests that negative consequences of disciplinarian or authoritarian parenting styles may be addressed most effectively through policies aimed at reducing the hazards faced by low-income children and feared by low-income parents, rather than by attempting to change parenting styles directly. In doing so, it provides a theoretical framework to explain reductions in harsh parental discipline and parental monitoring associated with assignment

to safe neighborhoods through court-ordered desegregation (Fauth, Leventhal and Brooks-Gunn 2007). Second, this paper provides new empirical evidence on the relationship between parental discipline and children's environment. While a large body of work has found associations between dangerous neighborhoods and harsh parental discipline (Pinderhughes, Foster and Jones 2001; Gayles et al. 2009; Gonzales et al. 2011; Anderson and Zhang 2010) and parental monitoring (Jones et al. 2005; Vieno et al. 2010; Cuellar, Jones and Sterrett 2015), this paper is the first to examine the relationship between school safety and parenting. Because dangerous schools are less likely to directly affect parents' stress and social isolation than are dangerous neighborhoods (Ceballo and McLoyd 2002), the relationship between parenting and school safety is more likely than that between parenting and neighborhood safety to be driven by parent's concern for their children. By finding evidence that parents discipline more and more harshly when their children attend dangerous schools, I lend support to the idea that disciplinarian parenting is strategic. Furthermore, because schools are under greater control of policy-makers than are other aspects of childrens' environment, understanding the effect of school safety on parents and children is of particular policy importance.

The rest of this paper will proceed as follows: Section 3.2 will develop a model of parent-child interactions in the presence of attractive hazards. Section 3.3 will describe the data used in the empirical section of this paper, Section 3.4 will describe the paper's empirical strategy, Section 3.5 will present primary evidence on the effect of school safety on parental discipline, Section 3.6 will

examine heterogeneous effects of school safety by race and poverty, and Section 3.7 will conclude.

Section 3.2: Model

I construct a model of interaction between altruistic parents and short-sighted children. In each of two periods of childhood, children decide how much trouble to get into. Because children are short-sighted, with beta-delta style preferences, they fail to fully account for the long-term ramifications of their actions. As a result, children misbehave more than is in their long-term interests. Prior to each period of childhood, parents modify children's choices in two ways: by investing in their children's decision-making skills (leading them to more fully account for the long-term effects of their decisions), and by punishing them for misbehavior. After childhood, children receive an adulthood payoff that is lowered by childhood misbehavior and increased by their decision-making ability. I will first describe the general form of this model and then derive a few comparative statics used to motivate the empirical work in the paper.

3.2.1: Model Set-Up

In each of two periods of childhood, children's utility $u_t(X_t, D_t)$ is given by the following:

$$u_t(X_t, D_t) = a_t m(D_t) - (X_t)\left(D_t - \widehat{D}_t\right) \quad (1)$$

Where m() and γ()are concave, increasing functions with m'(0) and $\gamma'(0) = \infty$. D_t represents the amount of delinquent behavior chosen by the child at time *t*.

X_t and $\widehat{D}_t$ are the choices of the parents: X_t is the amount of delinquency-focused discipline a parent administers, for misbehavior past the threshold level $\widehat{D}_t$. a_t represents the opportunities for delinquent behavior available to the child at time *t*. This utility function makes the following assumptions. Children enjoy getting into trouble, but also dislike being punished for delinquency. The first taste of delinquent behavior is irresistible, but each transgression is less satisfying than the last. Finally, small amounts of discipline are very effective, but discipline has diminishing marginal returns.

Decisions made during childhood affect a child's well-being in adulthood. I capture this by defining an adulthood payoff, determined by the child and parent's decisions in each period of childhood, and given by:

$$\nu(Z_1, Z_2, D_1, D_2) = \sigma_D(D_1, D_2) + \sigma_B(Z_1, Z_2) \qquad (2)$$

Where σ_B is a concave, increasing function with first derivative=∞ and σ_D is a convex, decreasing function with first derivative=0. Z_1 and Z_2 represent the investment parents make in children's decision-making and self-control in periods 1 and 2. Practically, these assumptions assert that while a little trouble doesn't hurt anyone, childhood misbehavior results in escalating negative consequences for children. Furthermore, the costs of delinquency in one period of childhood are increasing in delinquency in the other period of childhood. Likewise, investments in children's decision-making as adults have diminishing marginal returns, but investments in early childhood and late childhood are complementary (Heckman and Cunha 2007).

Children's ultimate well-being is a function both of their childhood happiness in each of two periods of childhood and of their adulthood payoff, and is given by the following:

$$U(X_1, X_2, Z_1, Z_2, D_1, D_2) = \sum_{i=1,2} \beta^{t-1} u_t(X_t, D_t) + \beta^2 \upsilon(Z_1, Z_2, D_1, D_2) \quad (3)$$

However, because children are short-sighted, they make decisions using quasi-hyperbolic discounting, discounting all future periods by $\eta(Z_1, Z_2) < 1$. As a result, children in period 1 maximize:

Max wrt. D_1:

$$\widetilde{U}_1(D_1) = u_1(X_1, D_1) + \eta_1(Z_1)\{\beta u_2(X_2, D_2) + \beta^2 \upsilon(Z_1, Z_2, D_1, D_2)\} \quad (4a)$$

And children in period 2 maximize:

Max wrt. D_2:

$$\widetilde{U}_2(D_2) = u_2(X_2, D_2) + \eta_2(Z_1, Z_2)\beta v_t(Z_1, Z_2, D_1, D_2) \quad (4b)$$

At the beginning of each period of childhood, parents modify their children's choices by choosing a level of punishment for delinquency, X_t, a level of unpunished delinquent behavior $\widehat{D}_t$, and a level of investment, Z_t, to maximize the child's expected lifetime utility. Following Weinberg (2001), I assume that parents are entirely altruistic, and have a fixed budget of parenting time which they allocate to the two tasks. Thus, in period 1, parents solve:

$$\text{Max wrt. } X_1, Z_1, \widehat{D}_1:$$

$$U_1(*) = \sum_{t \in 1,2} \beta^{t-1} u_t(X_t, D_t(X_t, Z_t)) + \beta^2 \upsilon(Z_1, Z_2, D_1, D_2) \quad (5a)$$

$$S.t. X_1 + Z_1 \leq 1$$

And in period 2, parents solve:

Max wrt. $X_2, Y_2, Z_2, \widehat{D}_2, \widehat{S}_2$:

$$U_2(*) = u_2(X_2, D_2(X_2, Z_2)) + \beta^2 v(Z_1, Z_2, D_1, D_2) \quad (5b)$$

$$S.t. X_2 + Z_2 \leq 1$$

I use this model to investigate the effect of the immediate-term attractiveness of delinquent behavior (denoted a_1 and a_2) on children's and parent's choices. To do so, I solve the model via backward induction, starting with the child's decision in period 2, who takes as given both her own choices in period 1 and her parent's choices in periods 1 and 2.

3.2.2: Child's period-2 decisions:

The child's period-2 first-order condition is:

$$\frac{\partial}{\partial D_2}: a_2 m'(D_2) - \gamma_x(X_2) + \eta_2 \beta \sigma_{2D}(D_1, D_2) = 0 \quad (6)$$

In order to calculate optimal parenting strategies, I first need to understand how children's decisions are affected by parenting decisions. To do this, I calculate a few comparative statics—particularly, the effect of delinquency-focused discipline on delinquency $\left(\frac{\partial D_2}{\partial X_2}\right)$ and the effect of parental investment on delinquency $\left(\frac{\partial D_2}{\partial Z_2}\right)$. Using the implicit function theorem, I find that:

$$\frac{\partial D_2}{\partial X_2} = \frac{\gamma\prime(X_2)}{a_2 m''(D_2) + \eta_2 \beta \sigma_{22D}(D_1, D_2)} \quad (7)$$

Thus, the magnitude of the effect of punishment on delinquency is an increasing function of the marginal effectiveness of punishment $\gamma'(X_2)$ and a decreasing function of the second derivative of the child's objective function with respect to

delinquency in the absence of punishment. Because $\gamma_x'(X_2) > 0$ and $a_2 m''(D_2) + \eta_2 \beta \sigma_{22D}(D_1, D_2) < 0$, this expression is negative—i.e. punishing delinquency lowers delinquency.

Likewise, investment in children's decision-making lowers delinquency:

$$\frac{\partial D_2}{\partial Z_2} = \frac{-\beta \eta_{2,z_2}(Z_1, Z_2)\sigma_{2D}(D_1, D_2)}{a_2 m''(D_2) + \eta_2 \beta \sigma_{22D}(D_1, D_2)} < 0 \quad (8)$$

The effect of investment on children's delinquency is an increasing function of the effect of investment on children's effective period-two time preferences times the marginal cost of period-two delinquent behavior on the adulthood payoff, and a decreasing function of the second derivative of the child's objective function with respect to delinquency in the absence of punishment. This is again negative, because $\sigma_{2D}(D_1, D_2) < 0$.

3.2.3: Parent's period-two decisions

As shown in equation 5b, parents choose X_2, Z_2, and $\hat{S}_2$ to maximize their children's utility, subject to a time constraint, and taking into account both the direct effect of each parenting input on their children's well-being and the effect on their children's period-two decisions. Because children's period-two decisions are deterministic given parent's period-two decisions, parents can choose $\hat{D}_2 = D_2$ for any given values of X_2 and Z_2. In other words, parents in this model influence their children through the threat of punishment, but target those

threats to avoid having to administer any actual punishment[18]. Thus, the direct effect of X_2 on children's utility can be set to zero. As a result, parents' first-order conditions are given by:

$$\frac{\partial}{\partial X_2}: \frac{\partial U}{\partial D_2} * \frac{\partial D_2}{\partial X_2} = \lambda_2 \tag{9a}$$

$$\frac{\partial}{\partial Z_2}: \frac{\partial \eta_2 \beta}{\partial Z_2} * \left(\frac{\partial U}{\partial D_2} * \frac{\partial D_2}{\partial \widetilde{\beta}_2} + \frac{\partial U}{\partial S_2} * \frac{\partial S_2}{\partial \widetilde{\beta}_2} + \frac{\partial U}{\partial \widetilde{\beta}_2}\right) = \lambda_2 \tag{9b}$$

From equation 3, $\frac{\partial U}{\partial X_2}$ is given by:

$$\frac{\partial U}{\partial D_2} = a_t m'(D_t) - \gamma(X_t) + \beta \sigma_{2D}(D_1, D_2) \tag{10a}$$

By subtracting the child's period 2 first-order condition (equation 6a), which is equal to zero, I find:

$$\frac{\partial U}{\partial D_2} = (\beta - \eta_2 \beta)\, \sigma_{2D}(D_1, D_2) \tag{10b}$$

Thus, the marginal effect of delinquency on a child's long-term wellbeing is the difference between the true value of the future and the child's consideration of the future, multiplied by the long-term marginal cost of delinquency.

Substituting equations 7 and 10b into equation 9a gives:

$$\frac{(\beta - \eta_2 \beta)\, \sigma_{2D}(D_1, D_2) \gamma_x'(X_2)}{a_2 m''(D_2) + \eta_2 \beta \sigma_{22D}(D_1, D_2)} = \lambda_2 \tag{11a}$$

Thus, delinquency-focused discipline has a positive return on child welfare because it is effective at reducing delinquency. The marginal utility of

[18] Allowing parents to discipline only after a certain level of delinquency is necessary to conclude that greater opportunities for delinquency will result in stricter punishment. If parents must punish all delinquent behavior, an increase in a child's delinquency increases the cost to the child's well-being of a strict punishment regime by necessitating more punishment. However, it is not necessary to entirely eliminate utility costs of discipline to children, so long as parents have some ability to "choose their battles." For instance, if an effective threat of punishment involves carrying out some actual punishment, to prove to children that their parents "mean it," children's utility may be directly decreasing in X_t and Y_t, but that direct utility effect will not depend directly on the child's delinquency and study effort.

delinquency-focused discipline is determined by the child's short-sightedness, the marginal cost of delinquent behavior in adulthood, and the effectiveness of discipline.

Likewise, substituting equations 8 and 10b into equation 9b gives:

$$\frac{(\beta-\eta_2\beta)\beta\eta_{2,z_2}(Z_1,Z_2)\sigma_{2D}{}^2(D_1,D_2)}{a_2m''(D_2)+\eta_2\beta\sigma_{22D}(D_1,D_2)}+\beta\,\sigma_{2B}(Z_1,Z_2)\ =\lambda_2 \qquad (11b)$$

I.e. investment in children's decision making gives a positive return on child welfare both because it is effective at reducing discipline and because it provides direct benefits to children when they enter adulthood. Furthermore, discipline and investment have the same marginal cost, the marginal benefit of the two investments are identical, and equal to the utility value of parents' time.

3.2.4: Comparative Statics on A_2

From here, I ask how a change in the appeal of delinquent behavior, denoted as a_2, will affect parents' choice of X_2, Y_2, and Z_2. I consider a change in a_2 that occurs between periods 1 and 2 and is not anticipated by parents or children. As a result, I treat D_1 and Z_1 as constants. Thus, an increase in a_2 will affect parents' choices both by entering into their first-order conditions (9a and 9b) directly and by changing children's choice of D_1. I first apply the implicit function theorem to 6a to calculate the effect of a_2 on D_2:

$$\frac{\partial D_2}{\partial a_2}=\frac{-m'(D_2)}{a_2m''(D_2)+\eta_2\beta\sigma_{22D}(D_1,D_2)}>0 \qquad (12)$$

Because $\frac{\partial^2 U}{\partial^2 D_2}<0$ (equation 10b), an increase in a_2 increases the marginal disutility of delinquency. As a result, because the returns to X_2 and Z_2 are

increasing both in a_2 directly and in D_2, an increase in a_2 will increase the return to both delinquency-preventing punishment X_2 and to investment Z_2. However, because an increase in either X_2 or Z_2 will also increase the value of parenting time λ_2, only one of these parenting can increase in a_2. I will show that the effect of a_2 on delinquency-focused discipline X_2 is unambiguously positive.

To see this, recall from 10a and 10b that the marginal utility of delinquency-suppressing discipline and investment in children's decision-making is identical. Because investment in a child has benefits beyond reducing delinquency, equilibrium levels of punishment and investment must be such that the marginal effect of punishment on delinquency is greater than the marginal effect of investment on delinquency. Formally, $\frac{\partial U}{\partial D_2} * \frac{\partial D_2}{\partial X_2} = \frac{\partial U}{\partial D_2} * \frac{\partial D_2}{\partial Z_2} + \frac{\partial U}{\partial Z_2}$. Because, $\frac{\partial U}{\partial S_2}$ and $\frac{\partial D_2}{\partial Z_2} * \frac{\partial U}{\partial Z_2}$ are both strictly positive, we know that $\frac{\partial U}{\partial D_2} * \frac{\partial D_2}{\partial X_2} > \frac{\partial U}{\partial D_2} * \frac{\partial D_2}{\partial Z_2}$, and thus that $\frac{\partial D_2}{\partial X_2} < \frac{\partial D_2}{\partial Z_2}$.

This finding is sufficient to ensure that $\frac{\partial D_2}{\partial a_2} > \frac{\partial Z_2}{\partial a_2}$, and thus that $\frac{\partial D_2}{\partial a_2} > 0$. Because a_2 has no effect on $\frac{\partial U}{\partial Z_2}$, it only affects parent decisions by increasing the marginal benefit of suppressing delinquency. The marginal effect of punishment on delinquency is greater than the marginal benefit of investment on delinquency, so punishment must increase more than investment with increased a_2. Because $X_2 + Z_2 = 1$, this requires that X_2 be increasing in a_2 and that Z_2 be decreasing in a_2.

We can therefore conclude that when children's environments become more hazardous, parents concerned only about controlling their children's immediate decisions will respond with increased discipline aimed at keeping their children out of trouble. However, because children raised in disadvantaged environments tend to remain in those environments throughout childhood, the more significant question is how persistently hazardous environments affect parenting. I will show that while the existence of immediate hazards will almost always lead to increased parental discipline, this is true only when immediate bad decisions have serious and irreversible consequences. If good behavior later in childhood can make up for bad behavior early in childhood, a more dangerous early childhood environment may not lead to increased discipline. To show this, I consider the child's and parent's period 1 decisions.

3.2.5: Child's Period 1 Decision:

Children in period 1 maximize $\widetilde{U}_1(D_1)$, as shown in equation 4a. Childrens' choice of D_1 has three effects: directly affecting immediate utility $u_1(D_1)$, affecting adult utility $v_3(Z_1, Z_2, D_1, D_2)$ and affecting parenting decisions in period 2 X_2 and Z_2. This is reflected in the child's period-1 first-order conditions:

$$\frac{\partial}{\partial D_1}: a_1 m'(D_1) - \gamma(X_1) + \eta_1 \left\{\beta^2 \sigma_{1D}(D_1, D_2) + \beta \left[\frac{\partial U_2}{\partial X_2}\frac{\partial X_2}{\partial D_1} + \frac{\partial U_2}{\partial Z_2}\frac{\partial Z_2}{\partial D_1}\right]\right\} = 0 \qquad (13)$$

However, while D_1 does change X_2 and Z_2, the effect on utility is zero. To see this, recall that X_2 and Z_2 are chosen to maximize U_2, conditional on $X_2 + Z_2 = 1$. Thus, $\frac{\partial U_2}{\partial X_2} = \frac{\partial U_2}{\partial Z_2} = \lambda_2$. As a result, we can rewrite 13 to read:

$$\frac{\partial}{\partial D_1}: a_1 m'(D_1) - \gamma_x(X_1) + \eta_1 \left\{\beta^2 \sigma_{1D}(D_1, D_2) + \beta\lambda \left[\frac{\partial X_2}{\partial D_1} + \frac{\partial Z_2}{\partial D_1}\right]\right\} = 0$$

Because $X_2 + Z_2 = 1$, we know that $\frac{\partial X_2}{\partial D_1} + \frac{\partial Z_2}{\partial D_1} = 0$. Thus, equation 13 is equal to:

$$\frac{\partial}{\partial D_1}: a_1 m'(D_1) - \gamma_x(X_1) + \eta_1 \beta^2 \sigma_{1D}(D_1, D_2) = 0 \tag{14}$$

From here, it is clear that children in period 1 face the same tradeoff as they will in period 2, only with a lower weight placed on adulthood payoffs. This allows us to derive the following comparative statics on the effects of period-1 parenting on period-1 child decisions:

$$\frac{\partial D_1}{\partial X_1} = \frac{\gamma_x'(X_1)}{a_1 m''(D_1) + \eta_1 \beta^2 \sigma_{11D}(D_1, D_2)} < 0 \tag{15a}$$

$$\frac{\partial D_1}{\partial Z_1} = \frac{\beta \eta_1'(Z_1) \sigma_{1D}(D_1, D_2)}{a_1 m''(D_1) + \eta_1 \beta^2 \sigma_{11D}(D_1, D_2)} < 0 \tag{15b}$$

We can also follow the derivation of children's period-2 decisions to determine the effect of a_1 on D_1:

$$\frac{\partial D_1}{\partial a_1} = \frac{-m'(D_1)}{a_1 m''(D_1) + \eta_1 \beta^2 \sigma_{11D}(D_1, D_2)} > 0 \tag{16}$$

3.2.6: Parent's Period 1 Decision:

Because their decisions influence children's actions in period 1 and in period 2, parents in period 1 face a substantially different problem than they do in period

two[19]. In addition to the effect of Z_1 on children's period-two delinquency through its effect on η_2, both parenting variables affect period-two choices because children's period-1 decisions affect the payoffs of period-2 decisions. Parents thus optimize according to the first-order conditions:

$$\frac{\partial}{\partial X_1}: \frac{\partial U}{\partial D_1}\frac{\partial D_1}{\partial X_1} + \frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}\frac{\partial D_1}{\partial X_1} = \lambda_1 \tag{17a}$$

$$\frac{\partial}{\partial Z_1}: \frac{\partial U}{\partial D_1}\frac{\partial D_1}{\partial Z_1} + \frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}\frac{\partial D_1}{\partial Z_1} + \frac{\partial \eta_2}{\partial Z_1}\left(\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial \eta_2}\right) + \frac{\partial U}{\partial Z_1} = \lambda_1 \tag{17b}$$

3.2.7: Comparative statics of a_1

From here, I investigate whether increased opportunities for delinquency increase discipline even in early childhood, when focusing on discipline, rather than investment, will make later childhood more difficult. To address this question, note that combining equations 17a and 17b gives:

$$\left(\frac{\partial U}{\partial D_1} + \frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}\right)\frac{\partial D_1}{\partial X_1} = \left(\frac{\partial U}{\partial D_1} + \frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}\right)\frac{\partial D_1}{\partial Z_1} + \frac{\partial \eta_2}{\partial Z_1}\left(\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial \eta_2}\right) + \frac{\partial U}{\partial Z_1} \tag{18a}$$

Because $\frac{\partial \eta_2}{\partial Z_1}\left(\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial \eta_2}\right) + \frac{\partial U}{\partial Z_1} > 0$, this in turn indicates that $\left(\frac{\partial U}{\partial D_1} + \frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}\right)\frac{\partial D_1}{\partial X_1} > \left(\frac{\partial U}{\partial D_1} + \frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}\right)\frac{\partial D_1}{\partial Z_1}$. We know that $\frac{\partial U}{\partial D_1} + \frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1} < 0$, because D_1 chosen by the child is above the welfare maximizing level of D_1. As a result, the total derivative of U with respect to D_1 must be negative. Therefore, we can conclude that $\frac{\partial D_1}{\partial X_1} < \frac{\partial D_1}{\partial Z_1}$, meaning that an increase in X_1 will reduce D_1 more than

[19] Effects on parents' decisions in period 2 can be ignored by the same argument put forward in Section 2.5.

will an equal increase in Z_1. To consider the effect of a_1 on X_1 and Z_1, consider the effect of a_1 on 18a:

$$\frac{\partial}{\partial a_1}\left\{\left(\frac{\partial U}{\partial D_1}+\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}\right)\left(\frac{\partial D_1}{\partial X_1}-\frac{\partial D_1}{\partial Z_1}\right)\right\}=\frac{\partial}{\partial a_1}\left\{\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial \eta_2}\frac{\partial \eta_2}{\partial Z_1}+\frac{\partial U}{\partial Z_1}\right\} \tag{18b}$$

This becomes:

$$\frac{\partial}{\partial a_1}\left[\frac{\partial U}{\partial D_1}+\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}\right]\left(\frac{\partial D_1}{\partial X_1}-\frac{\partial D_1}{\partial Z_1}\right)+\left(\frac{\partial U}{\partial D_1}+\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}\right)\frac{\partial}{\partial a_1}\left[\frac{\partial D_1}{\partial X_1}-\frac{\partial D_1}{\partial Z_1}\right]=\frac{\partial \eta_2}{\partial Z_1}\frac{\partial}{\partial a_1}\left[\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial \eta_2}\right]$$

Because $\frac{\partial D_1}{\partial X_1}$ and $\frac{\partial D_1}{\partial Z_1}$ are diminishing in X_1 and Z_1 respectively, we know that $\frac{\partial}{\partial a_1}>0$ implies that $\frac{\partial}{\partial a_1}[X_1-Z_1]>0$. We further know that the total derivative of U with respect to D_1, $\frac{\partial U}{\partial D_1}+\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}$ is negative, so:

$$\frac{\partial \eta_2}{\partial Z_1}\frac{\partial}{\partial a_1}\left[\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial \eta_2}\right]-\frac{\partial}{\partial a_1}\left[\frac{\partial U}{\partial D_1}+\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}\right]\left(\frac{\partial D_1}{\partial X_1}-\frac{\partial D_1}{\partial Z_1}\right)<0 \tag{19}$$

implies $\frac{\partial}{\partial a_1}[X_1-Z_1]>0$, and thus that $\frac{\partial X_1}{\partial a_1}>0$.

We know that the total derivative of U with respect to D_1, $\frac{\partial U}{\partial D_1}+\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial D_1}$, is decreasing in a_1, because the increase in D_1 with respect to a_1 $\frac{\partial D_1}{\partial a_1}$ is greater than the change that maximizes U_1. Because $\frac{\partial D_1}{\partial X_1}-\frac{\partial D_1}{\partial Z_1}<0$, the second term of expression 19 is positive.

In addition,

$$\frac{\partial \eta_2}{\partial Z_1}\frac{\partial}{\partial a_1}\left[\frac{\partial U}{\partial D_2}\frac{\partial D_2}{\partial \eta_2}\right]=-\frac{\partial \eta_2}{\partial Z_1}(1-\eta_2)\frac{\sigma_{22D}\sigma_{2D}\sigma_{12D}*(\sigma_{12D}-1)+\sigma_{2D}\sigma_{12D}-\frac{\partial}{\partial a_1}\left[\frac{\partial^2 U}{\partial^2 D_2}\right]\sigma_{2D}}{\left(\frac{\partial^2 U}{\partial^2 D_2}\right)^2} \tag{20}$$

In cases where $\frac{\partial}{\partial a_1}\left[\frac{\partial^2 U}{\partial^2 D_2}\right]\geq 0$ and $\sigma_{12D}>1$ the value of expression 20 is positive, resulting in an ambiguous prediction over whether $\frac{\partial X_1}{\partial a_1}>\frac{\partial Z_1}{\partial a_1}$ and thus whether

$\frac{\partial X_1}{\partial a_1} > 0$. However, we can determine that $\frac{\partial X_1}{\partial a_1} > 0$ in cases where $\frac{dU}{dD_1}$ is large, and where $\sigma_{12D}(D_1, D_2)$ is small.

Essentially, I find that parents will increase their punishment of young children when their children have more tempting opportunities to get in trouble, provided that young children's misbehavior has unavoidable future bad consequences. On the other hand, if young children's misbehavior is only costly if the misbehavior continues into adolescence, parents will increase investments in their children's decision-making skills in order to more effectively reduce adolescent misbehavior, rather than punishing young children. To see this, consider two extreme forms of the expression $\sigma_D(D_1, D_2)$. First, suppose that $\sigma_D(D_1, D_2) = \min(D_1, D_2)$, and D_1=D_2 at the current value of a_1. In this case, if $\frac{\partial D_1}{\partial a_1} > \frac{\partial D_2}{\partial a_1}$, $\frac{\partial U}{\partial D_1} = 0$, and the parent will choose to reduce period-2 delinquency through investment in period 1, rather than reduce period-1 delinquency through punishment in period 1. On the other hand, suppose that $\sigma_D(D_1, D_2) = D_1 + D_2$. In this case, $\sigma_{12D}(D_1, D_2) = 0$, and increased a_1 has no effect on children's choice of D_2 or on the welfare consequences of that choice. In this case, $\frac{\partial D_1}{\partial X_1} < \frac{\partial D_1}{\partial Z_1}$ is sufficient to ensure that $\frac{\partial X_1}{\partial a_1} > 0$. In other words, parents will discipline their children more aggressively in circumstances when their children have the opportunity to make significant, consequential errors, and when the consequences of those errors are not fully mitigated by learning from their mistakes.

3.2.8: Model Conclusion

This model has a few important implications. First, it suggests that when children grow up in environments where they might make consequential mistakes, parents will divert attention and effort towards keeping them from making those mistakes, at the expense of other kinds of parenting effort. In particular, the model unambiguously predicts that parents whose children face serious opportunities for and costs of misbehavior will divert effort away from pushing children to study. Because poor parents and parents from disadvantaged minority backgrounds raise children in environments that they describe as less safe (Pew Research Center 2015), this mechanism may explain why low income and Black parents punish their children more than do higher income, White, and Hispanic parents (Berlin et. al. 2009).

This model differs from others, particularly Cobb-Clark Salamaca and Zhu (2016), Weinberg (2001) and Cunha (2015) in that it attributes differences in discipline across families to differences in the external threats faced by children, rather than to differences in the resources, beliefs and characteristics of parents themselves. As a result, it predicts that parenting behavior can be altered by changing the environment that children inhabit without changing anything about parents themselves. Specifically, it predicts that, under most circumstances, children raised in environments where misbehavior is more attractive and/or carries greater risks will face more discipline from their parents, and that that discipline will be focused on preventing misbehavior. As

a result, it predicts that a change in the opportunities and costs of misbehavior during childhood will lead parents to increase their use of discipline. In the rest of this paper, I test these implications.

Section 3.3: Data and Measurement of Key Variables

In order to test this model empirically, I use the Los Angeles Family and Neighborhood Survey (LA.FANS)—a household-level panel survey conducted by the RAND Corporation in 2000 and in 2006. The LA.FANS is a survey of 2275 families with children from within Los Angeles County, selected at random from 65 census tracts. These tracts were selected as a stratified random sample from the over 1,800 tracts in Los Angeles County, with 20 tracts chosen from the poorest 10% of Los Angeles County tracts, 20 from tracts in the 60th-89th percentile of poverty, and 25 from tracts in the 59th percentile and below of poverty. Because the survey was designed to facilitate understanding of the effects of local geography on a variety of economic, sociological and health measures, the survey includes detailed information on where sample members live, work, go to school, and conduct other daily activities. In particular, the data include the census block group of residence for each household, at baseline and follow-up, as well as the name of the school attended by each of two randomly selected children in the household.

In addition to this detailed geographic information, The LA.FANS contains detailed information about parenting, adult behavior and well-being and child behavior and well-being, both from the parent's perspective and from the child's

perspective, for children above the age of eight. In each survey year, the adult member of the household self-identified as the "primary parent" is asked an array of questions about their child's personality, behavior, and development, as well as an array of questions about the parent's interactions with the child. In addition, each adult member of the household is asked questions about their own physical and mental well-being, their perceptions of their neighborhood, and their weekly activities. Finally, each child over the age of eight is asked a series of questions about their behavior at school and at home, their emotions, and their relationship with their parent. Each child is also given a series of cognitive and non-cognitive assessments.

The baseline survey includes 2275 families with children, each of whom are asked detailed questions about up to two children. The survey includes information on 3156 children, ranging in age from 0 to 17. The follow-up survey obtained responses from 1420 of the original 2275 families (62%), and 1133 of the original 3156 children (36%). Children were ineligible for the follow-up survey if they no longer resided in Los Angeles County (11%) or if they were over age 18 at the time of the follow-up survey (30%). In addition, 24% of respondent children refused to be interviewed for the follow-up survey (Peterson et. al. 2012).

Because I am interested in understanding the effect of school safety on parenting, I restrict this sample further to children who were attending school at the time of each survey. In addition, because I rely on the report of other children attending a child's school to determine the school's safety, I drop 20% of school-aged children at the baseline survey and 30% at the follow-up survey who are

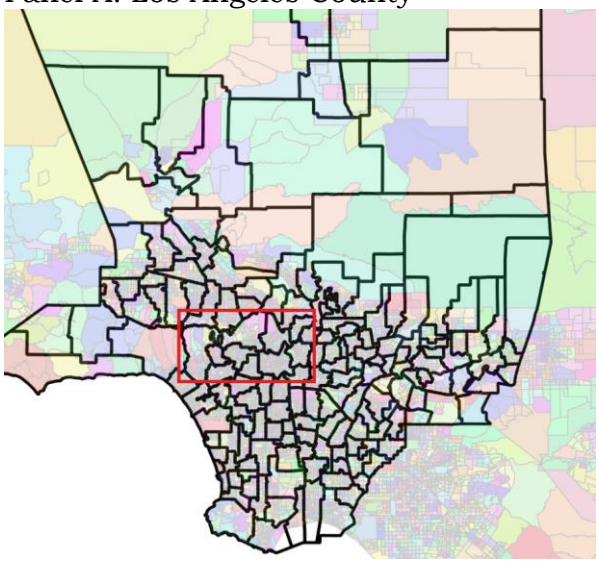
the only attendees of their school included in the data. These restrictions leave 1468 children in the baseline survey and 532 children in the follow-up survey who are included in cross-sectional analyses. In panel analyses, I restrict the sample to only those children who was expected to attend a school at the time of the follow-up survey that had at least one other child attending at both the baseline and follow-up surveys. This restriction leaves 329 children.

In addition to the LA.FANS data, I use a set of Los Angeles County school boundaries compiled into GIS files (Campbell 2002) to determine the elementary, middle, and high schools for which each child in the data is zoned. These boundary files were compiled and digitized by the Los Angeles County Emergency Operations Center in 2002 by collecting attendance area information from each school district in Los Angeles County. Because these files were collected in 2002, they represent school district boundaries in between the baseline and follow-up survey waves. I assign families to schools by determining which elementary, middle, and high school covers the largest fraction of each household's census block group of residence. Because census block groups contain an average of 1500 people, Los Angeles County block groups typically cover a fairly small area, with a median area of 1.3 square miles, the average census block group has 95% of its area within the boundaries of its selected High School, 93% within the boundaries of its selected middle school, and 81% within the boundaries of its selected elementary school. Figure 3.1 shows the relationship between these five geographies: high School, middle school and elementary school attendance areas, census tracts, and census block groups, for an area of Northeastern Los

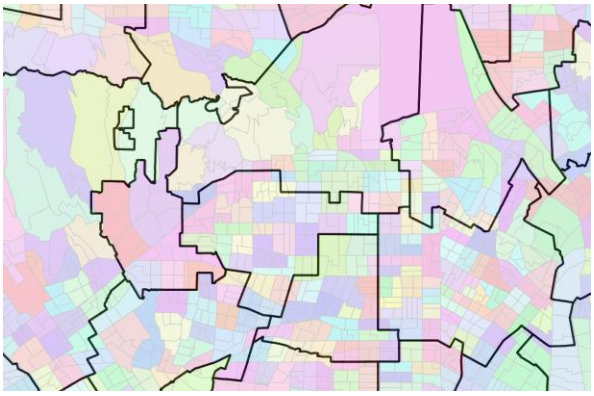
Angeles. While elementary school attendance areas are smaller than middle school and high school attendance areas, census block groups are very small relative to all school boundaries.

Figure 3.1: Geography

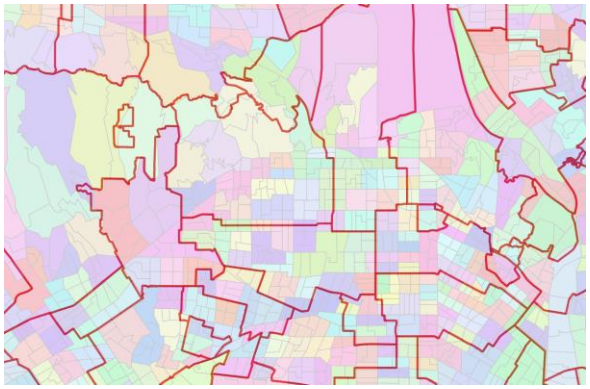
Panel A: Los Angeles County



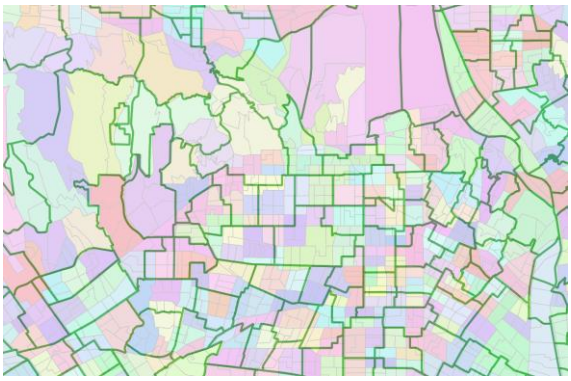
Panel B: High School Attendance Areas



Panel C: Middle School Attendance Areas



Panel C: Elementary School Attendance Areas



Notes: This figure shows the relative size of five geographies used in this analysis: High School, Middle School, and Elementary School attendance areas, Census Tracts, and Census Block Groups. The smallest divisions in each diagram represent census block groups. Block groups have been grouped and shaded to represent tracts. Panels A and B show High School boundaries (in black), Panel C shows Middle School attendance areas (in Red) and Panel D shows Elementary School attendance areas (in green). Panels B-D show the area of Northeast Los Angeles highlighted in the red rectangle in Panel A.

3.3.1: Measurement of Key Variables

A key challenge in this work is constructing meaningful measures of parenting, child behavior and development, and other social and emotional indicators from the survey questions available in the LA.FANS. With a few exceptions, parents and children are asked many different questions related to measures of interest in this research, such as what approach parents take to disciplining their children, how children perform in school, how much time parents spend investing in their children, and how much warmth is included in parents' relationships with their children. I make use of these multiple measures by using common factor analysis[20] to create summary measures from similar item questions.

To do this, I first categorized questions from each wave's parent survey, adult survey, and child survey by topic area, and then perform a multi-factor exploratory factor analysis on the variables included in each topic area. I then use the factor loadings on these multi-factor models to inform the number of separate indicators to construct and the inclusion of specific variables in each indicator. Finally, I used single-factor factor analytic results on these variables

[20] Common factor analysis is a technique used to estimate unobserved variables, referred to as "factors," from a number of different measures. The method uses maximum likelihood estimation to solve a system of linear regression equations, where each of N factors are independent variables and each measure is a dependent variable. The estimation selects both a set of values for the N factor (expressed as a linear combination of the dependent variables) and the regression coefficient on each factor in each regression. Because this estimation is under-identified when the number of factors is greater than 1, a specific solution, or "rotation" is selected to differentiate factors. In this analysis, I use a "promax" factor rotation, which allows factors to be correlated with each other and minimizes the regression coefficients of each dependent variable on factors other than the highest coefficient.

to construct scales, weighting each item by the fraction of its variation that is "common" to other items in the scale.

The key variables in this analysis are parental discipline and school safety. I construct two measures of each. The first measure of parental discipline is an index measuring how often a parent reports having spanked the child, grounded the child, sent the child to their room, or took away the child's privileges in the past week, with responses to each question weighted by their correlation to the other questions. I refer to this measure as "Actual Discipline." The second measure of parental discipline is an index capturing a parent's reported likelihood to spank their child, ground their child for more than 30 minutes, or scold their child if their child had a temper tantrum. I refer to this measure as "Intended Harsh Discipline." Each of these measures is constructed as a weighted average of a series of survey items, identified through common factor analysis as having a high level of covariance. I report the items included in these measures, the weights given to each item, and the correlation of the item to the final measures in Appendix Table D.1. Appendix Table D.1 also includes the items, weights, and correlations for three constructed controls capturing adult's perception of their neighborhood, which are used as controls.

I also capture two measures of school safety and one measure of school quality. Because a student's subjective report of their school's safety is likely to be confounded by the student's characteristics, all measures of school safety are calculated from the average survey responses of students in the LA.FANS data who attend a school, not counting the student for whom the measure is being

constructed. The first measure of school safety is calculated from the response of students to two questions: "Do other students misbehave" (reversed) and "Do you feel safe at this school?" I refer to this measure as "Reported School Safety." The second measure of school safety is calculated from student's report of engaging in various forms of misbehavior. For children above the age of 12, I use children's report of skipping school, misbehaving at school, getting in trouble with teachers, running away from home, smoking cigarettes, drinking alcohol, using marijuana, using drugs other than marijuana, selling drugs, having sex, getting pregnant/getting someone pregnant, belonging to a gang, carrying a gun, and being arrested by the police. For children under the age of 12, I use children's report of skipping school. I refer to this measure as "Reported School disorder." Finally, I include a measure of the quality of schools, other than safety, as a control. This control is defined by constructing an index from student's report of whether teachers are good, treat students fairly, and care about their students at their school. I refer to this control as "School Quality." Appendix Table D.2 reports the survey items included in each of these three measures, the weight given to each item when constructing the measures, and the correlation of the survey item to the final measure.

In order to determine the relationship between parenting measures constructed for this analysis, I examine the correlations between all 10 primary and secondary parenting measures constructed. These correlations are reported in Appendix Table D.3. As shown in Column 1, actual discipline is positively correlated with intended harsh discipline, but with a correlation of only 0.12.

This suggests that these two primary measures of discipline capture different aspects of parental discipline. Unsurprisingly, both actual discipline and intended harsh discipline are positively correlated with children's report of parents' monitoring, and negatively correlated to children's report of parents' warmth.

I also examine the correlations between measures of child behavior constructed for this analysis, as well as the correlations between child behavior and parenting behavior. These correlations are reported in Appendix Table D.4. Consistent with the separation of youth study and youth delinquency in my model, children's report of study effort is uncorrelated with their report of delinquency. It is also positively correlated with actual discipline, but uncorrelated with intended harsh discipline.

3.3.2: Characteristics of the Sample

The sample used for the primary analysis in this paper consists of children at the baseline and the follow-up who attend school with at least one other sample child and who have a parent who completed the parent survey. This provides a final sample of 2000 observations, taken from 1237 families. Table 3.1 shows the characteristics of this sample.

Table 3.1: Characteristics of Sample

Mean (SD)	Full Sample (1)	Hisp. (3)	White (4)	Black (5)	Asian (6)	Race Diff. (7)	Non-Poor (8)	Poor (9)	Pov. Diff (10)	Sample Size (11)
Prim.Parent HS Grad	0.46	0.27	0.88	0.76	0.73	***	0.58	0.33	***	1938
	(0.50)	(0.44)	(0.32)	(0.43)	(0.44)		(0.49)	(0.47)		
Prim. Parent Attended College	0.38	0.19	0.81	0.64	0.75	***	0.50	0.26	***	1961
	(0.49)	(0.39)	(0.39)	(0.48)	(0.43)		(0.50)	(0.44)		
Prim. Parent Graduated College	0.20	0.06	0.54	0.28	0.47	***	0.29	0.11	***	1894
	(0.40)	(0.24)	(0.50)	(0.45)	(0.50)		(0.45)	(0.31)		
Prim. Parent Born in US	0.35	0.17	0.80	0.92	0.21	***	0.44	0.26	***	1567
	(0.48)	(0.38)	(0.40)	(0.27)	(0.41)		(0.50)	(0.44)		
Household Assets (Thousands $)	24	13	50	26	45	***	33	10	***	1426
	(33)	(22)	(41)	(34)	(42)		(37)	(21)		
Household Earn. (Thousands $)	35	24	73	33	51	***	57	8	***	1783
	(41)	(25)	(63)	(33)	(41)		(44)	(7)		
Parents Married & Living Together	0.69	0.69	0.79	0.38	0.82	***	0.86	0.52	***	1947
	(0.46)	(0.46)	(0.41)	(0.49)	(0.38)		(0.35)	(0.50)		
Overwhelmed Parent	-0.02	-0.08	0.07	0.13	0.17	***	-0.07	0.03	**	1972
	(1.00)	(1.02)	(0.92)	(1.01)	(1.02)		(0.98)	(1.02)		
Parent Working	0.59	0.56	0.66	0.69	0.61	***	0.71	0.48	***	1992
	(0.49)	(0.50)	(0.47)	(0.46)	(0.49)		(0.45)	(0.50)		
Actual Discipline	-0.06	-0.02	-0.23	0.14	-0.25	***	-0.14	0.01	***	1158
	(0.94)	(0.95)	(0.79)	(1.10)	(0.91)		(0.87)	(0.99)		
Intended Harsh Discipline	0.02	-0.05	0.03	0.46	0.20	***	0.06	-0.02		1154
	(1.02)	(0.93)	(1.03)	(1.31)	(1.26)		(1.07)	(0.97)		
Parenting Warmth	0.03	0.02	0.18	-0.12	-0.04	**	0.08	-0.01	*	1516
	(0.98)	(0.97)	(0.91)	(1.12)	(1.03)		(0.92)	(1.04)		
Youth-Reported Delinquency	-0.03	-0.05	0.08	0.13	-0.21	***	-0.01	-0.05		1516
	(0.98)	(0.96)	(1.06)	(1.13)	(0.84)		(1.01)	(0.95)		
Youth-Reported Investment	0.05	0.02	0.09	0.03	0.19		0.11	-0.01	**	1523
	(1.03)	(1.05)	(1.00)	(0.97)	(0.99)		(1.06)	(1.00)		
Parent-Reported Youth Delinquency	-0.05	-0.03	-0.12	0.13	-0.15	**	-0.10	0.00	**	1999
	(0.99)	(0.98)	(0.93)	(1.12)	(1.00)		(0.95)	(1.02)		
Parent-Reported Youth Investment	0.04	0.04	0.13	-0.23	0.10	**	0.11	-0.02	**	1158
	(0.98)	(0.99)	(0.89)	(1.09)	(0.92)		(0.88)	(1.06)		
School Safety	0.00	-0.07	0.24	-0.09	0.15	***	0.03	-0.03		1998
	(1.00)	(1.02)	(0.84)	(1.11)	(0.89)		(0.95)	(1.05)		
School Quality	-0.00	-0.03	0.07	-0.03	0.12		0.07	-0.07	***	1998
	(1.00)	(0.98)	(1.01)	(1.00)	(1.09)		(0.96)	(1.03)		
School disorder	0.00	0.01	0.02	-0.03	-0.07		-0.03	0.03		2000
	(1.00)	(1.01)	(0.95)	(0.94)	(1.01)		(0.94)	(1.05)		
Observations	2000	1312	343	152	168		968	1032		

Notes: This table presents the average values of family characteristics in the analysis sample. Observations are at the child by year level.

Because the LA.FANS oversamples high-poverty census tracts, the sample examined in this analysis is disadvantaged, with fewer than 50% of parents completing high school and more than 50% with earnings below the poverty line. 2/3 of the sample are Hispanic, while 17% are non-Hispanic White, 8% are Black, and 8% are Asian, multiracial, or other.

In this sample, there are substantial differences in parenting, child behavior, and school and neighborhood quality by race and household income. Black parents and poor parents report using significantly more discipline than do other parents, with Black parents reporting use of discipline at rates 0.16 standard deviations above the mean for Hispanic parents and 0.37 standard deviations above the mean for White parents. Poor parents report using disciple at rates 0.15 standard deviations above the average for non-poor parents. Likewise, Black parents are far more likely than are non-Black parents to say that they would respond to misbehavior with harsh punishment, reporting that they would spank their child, scold their child, or send their child to their room for more than 30 minutes at rates 0.5 standard deviations above the mean for Hispanic parents, and 0.43 standard deviations above the mean for White parents. Poor and Black parents are also more likely than other parents to report that their children misbehave. However, Poor and Black parents do not misbehave at rates in great excess of their peers, with black children reporting higher rates of misbehavior higher than Hispanic and Asian children but comparable to White children. Poor parents report similar rates of misbehavior to non-poor children.

School and neighborhood quality also vary by race and income. On average, schools attended by Black and Hispanic children are rated as less safe than are schools attended by White and Asian parents. Schools attended by poor children are rated as lower quality than are children attended by non-poor children. While White parents are significantly less likely to interact regularly with their neighbors or be members of neighborhood organizations, White and Asian parents and children report feeling safer in their neighborhoods than do Black and Hispanic parents and children. Likewise, poor parents know more neighbors and participate in more neighborhood events than do non-poor parents, but poor parents and children rate their neighborhoods as more dangerous than do non-poor parents and children.

Section 3.4: Empirical Strategy

The primary empirical question of this paper is whether children raised in relatively hazardous environments receive stricter parenting than do similar children in less hazardous environments. Because I wish to isolate the effect on discipline resulting from changes in parents' perceptions of the risks facing children from changes in discipline resulting from parents' patience, I control for characteristics of families and of neighborhoods.

3.4.1: Relationship between school safety and parenting

To examine the persistent effect of disorderly schools on parenting, I estimate the following equation:

$$y_{kit}^{j,n} = \alpha + \beta S_{j,t} + \gamma n_t + \sigma X_{k,i,t} + \varepsilon_{kit} \quad (21)$$

Where $y_{kit}^{j,n}$ is parenting outcome y for child k in family i, attending school j in neighborhood n at time t. $S_{j,t}$ is the measured safety of the school j that the child attends at time t, n_t is a vector of fixed effects for neighborhood n at time t, and $X_{k,i,t}$ is a matrix of time-varying, family and child-level controls.

The key empirical challenge in understanding this relationship is that family disadvantage is positively associated with the use of harsh discipline and negatively associated with school quality. As a result, unless the relationship between family disadvantage and discipline is driven entirely by children's more dangerous environments, the correlation between school disorder and discipline will be more negative than is the causal relationship between school disorder and discipline. While I cannot entirely eliminate this potential bias, I address this concern through the use of controls for confounding family and neighborhood characteristics.

I control for several characteristics of families that are most likely to be related both to school quality and to parental discipline. This includes controls for family income, child age, and parent race, education and marital status, as well as parents' engagement in their neighborhood, consumption of alcohol, and report of depression. While these controls are not adequate to entirely control for

differences between families, the fact that the estimated relationship between school safety and parenting remains consistent with the inclusion of these controls suggests that the relationship is not driven primarily by economic and emotional differences between families.

An additional source of endogeneity comes from the relationship between neighborhoods and schools. Schools in safe and wealthy neighborhoods are likely to be safer and more controlled environments than are schools in low income and dangerous neighborhoods. In turn, neighborhood quality can affect parenting by affecting the social isolation of parents (Pinderhuges et. al. 2007), parents' stress and emotional reserves (Kohen et. al. 2008) and by affecting parents' economic opportunities and expected opportunities for kids. In addition, neighborhood is likely related to myriad unobserved personal characteristics of parents.

I address this in two ways. First, I include census tract by survey year fixed effects, in order to directly control for unobserved characteristics of neighborhoods and the families living in them. Because census tracts are fairly small within Los Angeles County, and because they only partially overlap with school boundaries (See Figure 3.1), tract-by-year-level fixed effects are likely to capture most aspects of neighborhoods that might be related to parenting. Additionally, I control directly for parents' report in each year of neighborhood quality.

3.4.2: Heterogeneous Effects

I examine two sources of heterogeneity in effect: whether a family is below the poverty line, and whether the primary parent is Black. I examine these sources of heterogeneity for several reasons. First, an extensive literature demonstrates that Black children are punished in schools at higher rates than are children from other ethnic backgrounds (Wallace et al 2008; Tobin 2011), and are more likely to be arrested (Tapia 2010) and to be victims of crime (Snyder and Sickmund 2006) during childhood. As a result, a dangerous school environment may pose a greater risk to black children than to children of other racial backgrounds. Likewise, poor children are likely to be at greater disadvantage in similar school environments. Second, an extensive literature examining the relationship between neighborhood context and parenting finds that low income and Black parents change parenting practices in response to dangerous and disadvantaged neighborhoods to a greater extent than do higher income and non-Black parents (Zhang and Anderson 2010). I examine sources of heterogeneity in the cross-sectional relationship between school safety and parental discipline using the following specification:

$$y_{kit}^{j,n} = \alpha + \beta S_{j,t} + \vartheta Z_{j,t} S_{j,t} + \theta Z_{j,t} + \gamma n_t + \sigma X_{k,i,t} + \varepsilon_{kit} \qquad (22)$$

Where $Z_{j,t}$ is the source of heterogeneity of interest—either race or poverty status.

Section 3.5: Results

As a first test of the validity of the school safety measures defined in Section 3.3.1, I examine the effect of school safety on child misbehavior. The hypothesized relationship between school safety and parental discipline flows through child misbehavior—requiring that dangerous schools either provide more opportunities for children to engage in delinquent behaviors or that they increase the costs of doing so. As shown in Table 3.2, I find that "Reported School disorder" is positively associated with youth misbehavior, even after controlling for family and neighborhood characteristics.

Table 3.2: School Disorder on Child-Reported Delinquency

Dep Var: Child-reported delinquency	(1)	(2)	(3)	(4)	(5)	(6)
Reported School disorder	0.20***	0.12***	0.11***	0.09**	0.15***	0.08
	(0.03)	(0.03)	(0.03)	(0.04)	(0.05)	(0.06)
Child < 8				0.00***	0.00***	0.00***
				(0.00)	(0.00)	(0.00)
Child Age				0.05	0.03	0.12**
				(0.03)	(0.04)	(0.05)
Household Earnings				0.00	0.00	0.00
				(0.00)	(0.00)	(0.00)
Parents Married & Living Together				-0.07	-0.12	-0.14
				(0.07)	(0.09)	(0.12)
Income relative to poverty				0.01	-0.04	-0.07
				(0.06)	(0.08)	(0.10)
Below poverty line				-0.12	-0.08	-0.14
				(0.08)	(0.10)	(0.14)
Below 200% of poverty line				-0.06	-0.05	-0.17
				(0.11)	(0.14)	(0.17)
Prim.Parent HS Grad				0.07	0.01	-0.04
				(0.10)	(0.11)	(0.16)
Prim. Parent Attended College				-0.14	-0.04	0.03
				(0.12)	(0.14)	(0.19)
Prim. Parent Graduated College				-0.19*	-0.16	-0.24
				(0.11)	(0.13)	(0.17)
Parent: Friends in Neighborhood					0.04	-0.00
					(0.05)	(0.06)
Parent: Participates in Neighborhood					0.07*	0.11**
					(0.04)	(0.04)
Parent: Feels safe in Neighborhood					0.03	0.05
					(0.05)	(0.06)
Parent: Drinks per week						-0.02
						(0.02)
Parent: 4 or more drinks at a time per week						-0.00
						(0.04)
Parent: reported depression						0.27
						(0.19)
Wave by Tract FE	No	Yes	Yes	Yes	Yes	Yes
School Quality Control	No	No	Yes	Yes	Yes	Yes
Race	No	No	No	Yes	Yes	Yes
Observations	1479	1479	1477	1263	815	576
Dependent Variable Mean	-0.0311	-0.0311	-0.0334	-0.0540	-0.0636	-0.0335

Notes: This table reports estimated coefficients from the regression described in equation (21). The dependent variable is an index of child misbehavior. The table reports a regression of child misbehavior on other students' report of misbehavior at the child's school. Column (2) adds controls for census tract by survey wave. Column (3) adds a control for other students' report of school quality. Significance levels are: * 10%, ** 5%, *** 1%.

As shown in column (1) of Table 3.2, a 1 standard-deviation increase in reported school disorder is associated with a 0.2 standard-deviation increase in child misbehavior. As shown in Column (2), the introduction of wave by census tract fixed-effects reduces this effect size to 0.12 standard deviations, indicating that children who live in census tracts with high levels of youth misbehavior also attend schools with high levels of youth misbehavior. The introduction of controls for school quality (Column 3), child age parent race, earnings and poverty status, and parent education (column 4) further reduces the estimated effect size to 0.09 standard deviations. Further controls for parent's perception of neighborhood quality (column 5) increase the estimated effect size to 0.15 standard deviations, but reduce the sample by more than 400 observations due to high non-response rates. Controlling for parent's alcohol consumption and reported depression reduces the effect size to 0.08, but also reduces the usable sample to only 576.

These findings suggest that the measure of "reported school disorder" used in this analysis is related to the opportunities and attractiveness of misbehavior resulting from school environment. While neighborhood is, unsurprisingly, an important mediator of this relationship, the fact that controls for parent and child characteristics do not significantly alter the relationship between child delinquency and reported school disorder suggests that this relationship does not purely reflect student disadvantage. Because the positive relationship between school disorder and youth delinquency suggests that school disorder meaningfully measures the attractiveness of or opportunity for misbehavior at

school, I use this measure as the primary measure of school safety. Surprisingly, the relationship between student's direct report of school safety—referred to above as "Reported School Safety" and youth delinquency is also positive, suggesting that this measure of safety may capture the cost, rather than the attractiveness or opportunity, for misbehavior at school. This relationship is reported in Appendix Table D.5. As shown in appendix tables A.5 and A.6, the measure of "Reported School Safety" is not related to actual discipline or intended harsh discipline.

Section 3.5.1: Relationship between School Safety and Discipline

I estimate the relationship between my primary school safety measure, referred to as "reported school disorder," and two measures of parental discipline: "actual discipline" and "intended harsh discipline." These measures are described in Section 3.3.2, and are defined in table A.2. Table 3.3 shows the estimated cross-sectional relationship between "reported school disorder" and "intended harsh discipline," estimated using equation 22.

Table 3.3: School Disorder on Intended Harsh Discipline

Dep Var: Intended Harsh Discipline	(1)	(2)	(3)	(4)	(5)	(6)
Reported School disorder	0.05	0.09***	0.11***	0.10**	0.11**	0.08
	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.07)
Child < 8				0.07	0.09	-0.04
				(0.16)	(0.23)	(0.26)
Child Age				0.07*	0.11**	0.05
				(0.04)	(0.05)	(0.06)
Household Earnings				-0.00	0.00	0.00
				(0.00)	(0.00)	(0.00)
Parents Married & Living Together				0.09	-0.02	0.03
				(0.09)	(0.12)	(0.14)
Income relative to poverty				-0.00	0.01	-0.04
				(0.07)	(0.08)	(0.10)
Below poverty line				-0.03	0.06	0.26
				(0.11)	(0.14)	(0.18)
Below 200% of poverty line				-0.13	-0.21	-0.47**
				(0.16)	(0.18)	(0.23)
Prim.Parent HS Grad				-0.15	-0.33**	-0.40**
				(0.13)	(0.15)	(0.18)
Prim. Parent Attended College				0.08	0.23	0.31
				(0.15)	(0.17)	(0.20)
Prim. Parent Graduated College				-0.26*	-0.28	-0.48**
				(0.15)	(0.22)	(0.23)
Parent: Friends in Neighborhood					0.05	0.03
					(0.06)	(0.07)
Parent: Participates in Neighborhood					0.00	0.08
					(0.06)	(0.05)
Parent: Feels safe in Neighborhood					0.07	0.02
					(0.07)	(0.08)
Parent: Drinks per week						-0.05**
						(0.02)
Parent: 4 or more drinks at a time per week						0.05
						(0.05)
Parent: reported depression						-0.11
						(0.21)
Wave by Tract FE	No	Yes	Yes	Yes	Yes	Yes
School Quality Control	No	No	Yes	Yes	Yes	Yes
Race	No	No	No	Yes	Yes	Yes
Observations	1120	1120	1119	943	605	434
Dependent Variable Mean	0.0221	0.0221	0.0225	0.0233	0.0005	-0.0002

Notes: This table reports estimated coefficients from the regression described in equation (21). The dependent variable is an index of harsh responses to hypothetical misbehavior. The table reports a regression of intended harsh discipline on average reported misbehavior at the child's school. Column (2) adds controls for census tract by survey wave. Column (3) adds a control for other students' report of school quality. Significance levels are: * 10%, ** 5%, *** 1%.

As shown in Table 3.3, parents of students who attend schools where their peers get into more trouble are more likely to say that they would hit their child, scold their child, or send their child to their room for more than 30 minutes in response to a temper tantrum. While the raw correlation between school disorder and intended harsh discipline suggests a statistically insignificant effect size of only 0.05 (column 1), the inclusion of census tract fixed effects increases the estimated effect size to 0.09 (column 2), with controls for school quality (column 3), child age, household earnings, family structure, and parent education (column 4), parent's satisfaction with their neighborhood (column 5) and parent alcohol consumption and depression (column 6) having very little effect on the estimated coefficient. These results support the hypothesis that parents discipline their children more harshly when children's environments contain more opportunities to make costly mistakes.

Surprisingly, however, I find no evidence of a relationship between school safety and the actual amount of discipline administered by parents. As shown in Table 3.4, there is no cross-sectional relationship between reported school disorder and actual discipline regardless of the inclusion or exclusion of controls.

Table 3.4: School Disorder on Actual Discipline

Dep. Var: Actual Discipline	(1)	(2)	(3)	(4)	(5)	(6)
Reported School disorder	-0.01	-0.03	-0.02	-0.01	0.01	-0.01
	(0.03)	(0.04)	(0.04)	(0.04)	(0.06)	(0.08)
Child < 8				-0.15	-0.04	-0.08
				(0.17)	(0.24)	(0.26)
Child Age				-0.11***	-0.12**	-0.11*
				(0.03)	(0.05)	(0.06)
Household Earnings				0.00*	0.00**	0.00***
				(0.00)	(0.00)	(0.00)
Parents Married & Living Together				0.14*	0.07	0.14
				(0.08)	(0.12)	(0.14)
Income relative to poverty				-0.09*	-0.12**	-0.17**
				(0.05)	(0.05)	(0.07)
Below poverty line				0.14	0.11	0.03
				(0.10)	(0.14)	(0.16)
Below 200% of poverty line				0.05	0.12	0.30
				(0.12)	(0.15)	(0.18)
Prim.Parent HS Grad				0.22*	0.16	0.05
				(0.12)	(0.16)	(0.18)
Prim. Parent Attended College				-0.06	-0.11	0.16
				(0.14)	(0.19)	(0.25)
Prim. Parent Graduated College				-0.09	0.03	0.18
				(0.11)	(0.17)	(0.20)
Parent: Friends in Neighborhood					-0.1*	-0.13**
					(0.06)	(0.06)
Parent: Participates in Neighborhood					0.04	0.04
					(0.05)	(0.05)
Parent: Feels safe in Neighborhood					0.11**	0.14**
					(0.06)	(0.07)
Parent: Drinks per week						0.00
						(0.02)
Parent: 4 or more drinks at a time per week						0.07
						(0.05)
Parent: reported depression						0.16
						(0.26)
Wave by Tract FE	No	Yes	Yes	Yes	Yes	Yes
School Quality Control	No	No	Yes	Yes	Yes	Yes
Race	No	No	No	Yes	Yes	Yes
Observations	1124	1124	1123	947	609	436
Dependent Variable Mean	-0.0677	-0.0677	-0.0673	-0.0712	-0.0639	-0.0717

Notes: This table reports estimated coefficients from the regression described in equation (21). The dependent variable is a standardized index of discipline types used in the past week. The table reports a regression of actual discipline on average reported misbehavior at the child's school. Column (2) adds controls for census tract by survey wave. Column (3) adds a control for other students' report of school quality. Significance levels are: * 10%, ** 5%, *** 1%.

As shown in columns 1, 2, and 3, specifications that use the full sample of children with both measured school disorder and measured parental discipline suggest that an effect size larger than 0.06 is not supported by the data. This finding is consistent with a literal reading of my model, which suggests that parents discipline more aggressively, but ignore more serious infractions, when children are in more hazardous environments. It may also be explained by the fact that, as shown in Appendix Table D.7, parents' report of their children's misbehavior is unrelated to school disorder. Thus, parents whose children attend disorderly schools may respond to misbehavior more harshly than do parents whose children attend safer schools, but may not be any more likely to detect misbehavior.

Section 3.6: Heterogeneity in Effect on Parental Discipline

As discussed in Section 3.4.3, the effect of school disorder on discipline may be especially large among Black parents and parents raising children in poverty because Black children and poor children are more likely than are other children to face serious consequences for misbehavior. I find strong evidence in favor of this hypothesis.

Table 3.5 shows heterogeneity by poverty status in the relationship between school disorder and discipline. As shown in columns 1-3, the relationship between school disorder and actual discipline is significantly more positive for poor parents than for non-poor parents. When controlling for family, school, and neighborhood characteristics (Column 3), I find that a 1 standard

deviation increase in school disorder is associated with a 0.10 standard deviation decrease in actual discipline for non-poor parents but a 0.11 standard deviation *increase* in actual discipline for poor parents. In contrast, the effect of school disorder on intended harsh disciplined does not appear to vary by family poverty, with a 1 standard deviation increase in reported school misbehavior associated with a 0.09 standard deviation increase in intended harsh discipline regardless of poverty status.

Table 3.5: Heterogeneity by Poverty

Dependent:	Actual Discipline			Intended Harsh Discipline		
	(1)	(5)	(6)	(1)	(1)	(1)
Reported School disorder	-0.10***	-0.10**	-0.10**	0.06	0.10*	0.09
	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)	(0.06)
Reported School disorder X Below Pov. Line	0.16***	0.20***	0.21***	-0.02	-0.03	-0.00
	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)
Below Pov. Line	0.15***	0.12	0.12	-0.07	-0.04	-0.05
	(0.06)	(0.09)	(0.09)	(0.06)	(0.10)	(0.10)
Child < 8		-0.05	-0.09		0.05	0.07
		(0.15)	(0.16)		(0.15)	(0.16)
Child Age		-0.09***	-0.09***		0.06*	0.05
		(0.03)	(0.03)		(0.03)	(0.03)
Household Earnings		0.00	0.00*		0.00	-0.00
		(0.00)	(0.00)		(0.00)	(0.00)
Parents Married & Living Together		0.13*	0.15*		0.05	0.05
		(0.08)	(0.08)		(0.09)	(0.09)
Income relative to poverty		-0.05	-0.07*		-0.03	-0.01
		(0.04)	(0.04)		(0.06)	(0.06)
Below 200% of poverty line		0.10	0.11		-0.07	-0.12
		(0.10)	(0.11)		(0.14)	(0.15)
Prim.Parent HS Grad			0.20*			-0.15
			(0.11)			(0.12)
Prim. Parent Attended College			-0.01			0.09
			(0.14)			(0.15)
Prim. Parent Graduated College			-0.11			-0.22
			(0.11)			(0.15)
Wave by Tract FE	No	Yes	Yes	No	Yes	Yes
Race, School Quality Controls	No	Yes	Yes	No	Yes	Yes
Observations	1124	1014	947	1120	1010	943
Dependent Variable Mean	-0.0677	-0.0620	-0.0712	0.0221	0.0226	0.0233

Notes: This table reports estimated coefficients from the regression described in equation (22). In columns (1)-(3), The dependent variable is a standardized index of discipline types used in the past week. In columns (4)-(6), the dependent variable is an index of harsh responses to hypothetical child misbehavior. The table reports a regression of these dependent variables on the average report of misbehavior at the child's school, and an interaction of reported misbehavior and poverty. Significance levels are: * 10%, ** 5%, *** 1%.

While these findings are consistent with the hypothesis that poor families are more seriously affected by hazardous school environments than are higher income families, it is surprising that the relationship between school disorder and discipline is negative for non-poor households. One possible explanation for this is differences in the selection of poor-and non-poor households into disorderly schools. Because higher income families have more choice of where to live and where to send their children to school than do poor families, higher income families who send their children to relatively hazardous schools may be either less concerned with childhood misbehavior or less aware of children's school environment than do families who send their children to safer schools. For similar reasons, these less-concerned parents may discipline children less than do more-concerned parents, introducing negative bias to the estimated relationship between school disorder and parenting among higher-income households. This bias may affect actual discipline more than intended harsh discipline because actual discipline is more likely to reflect parental engagement.

Similarly, I find that Black parents' discipline decisions are more strongly related to school disorder than are non-Black parents' discipline decisions. As shown in Columns 1-3 of Table 3.6, Black parents engage in more discipline when their children attend a disorderly school. When controlling for socioeconomic status, neighborhood, and school quality (Column 3), a 1 standard deviation increase in school disorder is associated with a 0.17 standard deviation increase in actual discipline by Black parents, relative to a 0.02 standard deviation increase in actual discipline by non-black parents. Likewise,

as shown in Columns 4-6, black parents' intended harsh discipline is also more strongly related to school disorder than are non-black parents' intended harsh discipline. When including the full set of controls (Column 3), a 1 standard deviation increase in school disorder is associated with a 0.24 standard deviation increase in intended harsh discipline for Black parents, but only a 0.08 standard deviation increase for non-Black parents.

Table 3.6: Heterogeneity by Race

Dependent:	Actual Discipline			Intended Harsh Discipline		
	(1)	(5)	(6)	(1)	(1)	(1)
Reported School disorder	-0.03 (0.03)	-0.01 (0.04)	0.02 (0.05)	0.04 (0.03)	0.08** (0.04)	0.08* (0.04)
Reported School disorder X Black	0.22 (0.17)	0.18 (0.17)	0.15 (0.14)	0.16 (0.16)	0.15 (0.20)	0.16 (0.20)
Black	0.23* (0.12)	0.28 (0.20)	0.34 (0.25)	0.48*** (0.14)	0.36 (0.25)	0.39 (0.27)
Child < 8		-0.07 (0.16)	0.03 (0.22)		0.05 (0.15)	0.07 (0.16)
Child Age		-0.09*** (0.03)	-0.09** (0.04)		0.06* (0.03)	0.05 (0.03)
Household Earnings		0.00** (0.00)	0.00*** (0.00)		0.00 (0.00)	-0.00 (0.00)
Parents Married & Living Together		0.14* (0.08)	0.09 (0.11)		0.06 (0.09)	0.06 (0.09)
Income relative to poverty		-0.08* (0.04)	-0.10** (0.05)		-0.03 (0.06)	-0.01 (0.06)
Below 100% of poverty line		0.11 (0.09)	0.11 (0.12)		-0.04 (0.10)	-0.05 (0.10)
Below 200% of poverty line		0.12 (0.11)	0.16 (0.14)		-0.06 (0.14)	-0.11 (0.15)
Prim.Parent HS Grad			0.18 (0.15)			-0.15 (0.12)
Prim. Parent Attended College			-0.09 (0.17)			0.10 (0.15)
Prim. Parent Graduated College			-0.00 (0.15)			-0.22 (0.15)
Wave by Tract FE	No	Yes	Yes	No	Yes	Yes
Race, School Quality Controls	No	Yes	Yes	No	Yes	Yes
Observations	1124	1014	947	1120	1010	943
Dependent Variable Mean	-0.0677	-0.0620	-0.0712	0.0221	0.0226	0.0233

Notes: This table reports estimated coefficients from the regression described in equation (23). In columns (1)-(3), The dependent variable is a standardized index of discipline types used in the past week. In columns (4)-(6), the dependent variable is an index of harsh responses to hypothetical child misbehavior. The table reports a regression of these dependent variables on the average report of misbehavior at the child's school, and an interaction of reported misbehavior and Black primary parent. Significance levels are: * 10%, ** 5%, *** 1%.

While differences in response to school disorder by race are not statistically significant, they are large in magnitude and consistent with discipline being employed most when the costs of child misbehavior are highest. Furthermore, it is important to note that Black households in our sample have average earnings—the average Black household earns $33,000/year, compared to $35,000/year for the sample as a whole. As a result, poverty and race are distinct dimensions of disadvantage. It is thus notable that the discipline styles of both poor and Black households are more strongly related than average to school disorder.

Section 3.7: Conclusion

In this paper, I construct a model of parent-child interactions in which altruistic parents modify the choices of their short-sighted children by punishing bad decisions and teaching their children to consider the future more fully. From this model, I reach the conclusion that parents raising children under hazardous conditions will put more of their effort toward punishing their children to keep them safe, and will thus have fewer resources left to either encourage children to study or to invest in children's non-cognitive abilities and decision-making skills. I find evidence consistent with this model in an examination of the effect of school safety on parenting in Los Angeles County from 2000-2006. I find that parents are more likely to describe harsh parenting strategies when their children attend hazardous schools. This effect is larger for poor and Black families, who likely face more hazardous environments and greater costs of

misbehavior than do other families. While average rates of reported frequency of discipline are not related to the safety of schools in the full sample, poor and black families use more discipline when their children attend less safe schools.

This model suggests that negative consequences of harsh or disciplinarian parenting practices might be most effectively remedied by working to make life less dangerous for disadvantaged children, rather than by directly intervening in parenting. Because discipline is focused on protecting children from the harmful outcomes of short-sighted decisions, such policies need to both reduce the hazards facing poor children—for instance by reducing crime, decreasing the availability of drugs, etc, but also by employing milder punishment for children who make mistakes. The importance and complexity of this issue highlights the need for research understanding the links between parenting and crime prevention, zero-tolerance policies, police brutality, and other policy interventions that affect the safety of disadvantaged children.

Works Cited

Chapter 1:

Acemoglu, D., Autor, D. H., & Lyle, D. (2004, 6). Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury,. *Journal of Political Economy, 112*, 497–551.

Addison, J. T., Ozturk, O. D., & Wang, S. (2015). The Occupational Feminization of Wages,. *GEMF Working Papers2015-12*.

Amuedo-Dorantes, C., Rica, D. L., & Sara. (2006). The Role of Segregation and Pay Structure on the Gender Wage Gap: Evidence from Matched Employer-Employee Data for Spain. *The B.E. Journal of Economic Analysis & Policy, 5*, 1–34.

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration,. *The Quarterly Journal of Economics, 118*, 1279–1333.

Autor, D., Figlio, D., Karbownik, K., & Melanie Wasserman, J. R. (2016, 5). School Quality and the Gender Gap in Educational Achievement,. *American Economic Review, American Economic Association, 106*, 289–95.

Baron, J. N., Hannan, M. T., & Ozgecan Kocak, G. H. (2002). Gender and the Organization-Building Process in Young, High-Tech Firms. In M. Guillen, R. Collins, & P. E. Marshall Meyer (Eds.). Russell Sage Foundation.

Basit Zafar, Y. Q. (2009). Do Faculty Role Models Influence Female Students. *Choice of Major, and Why?" Journal of Economic Literature*.

Basu, K. (2015). Discrimination as a Coordination Device : Markets and the Emergence of Identity. *Policy Research Working Paper;No, 7490*. Retrieved from https://openknowledge.worldbank.org/handle/10986/23454

Becker, G. S. (1985). Human Capital, Effort, and the Sexual Division of Labor. *Journal of Labor Economics, 3*, 33–58. Retrieved from www.jstor.org/stable/2534997.

Bergmann, B. R. (1974). Occupational Segregation, Wages and Profits When Employers Discriminate by Race or Sex. *Eastern Economic Journal, 1*, 103–110.

Bertrand, M., & Pan, J. (2013). The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. *American Economic Journal: Applied Economics, 5*, 32–64.

Blau, F. D., Brummond, P., Liu, & Albert. (2013). Yung-Hsu "Trends in Occupational Segregation by Gender 1970-2009: Adjusting for the Impact of Changes in the Occupational Coding System.". *Demography, 50*, 471–494.

Blau, F., & Kahn, L. (2016). The Gender Wage Gap: Extent, Trends, and Explanations. In *IZA Working Paper Series.*

Catanzarite, L. (n.d.). Race-Gender Composition and Occupational Pay Degradation. *Social Problems, 50*, 14–37. Retrieved from www.jstor.org/stable/10.1525/sp.2003.50.1.14.

Chen, A., Keller, B., Lotker, Z., Mathieu, C., Peleg, D., & Pignolet, Y.-A. (2015). Homophily and the Glass Ceiling Effect in Social Networks. *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS '15). ACM.* New York, NY, USA, 41-50. doi:DOI

Cohen, P. N., & Huffman, M. L. (n.d.). Occupational Segregation and the Devaluation of Women's Work across U.S. Labor Markets. *Social Forces, 81*, 881–908. Retrieved from www.jstor.org/stable/3598179.

Cotter, D., DeFiore, J., Hermsen, J., Kowalewski, B., & Vanneman, R. (1997). All Women Benefit: The Macro-Level Effect of Occupational Integration on Gender Earnings Equality. *American Sociological Review, 62*, 714–734. Retrieved from http://www.jstor.org/stable/2657356De

Currie, J., & Moretti, E. (2003, 11). Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings. *The Quarterly Journal of Economics, Volume, 118*, 1495–1532. Retrieved from https://doi.org/10.1162/003355303322552856

Deming, D. J. (2015). The Growing Importance of Social Skills in the Labor Market,. In *NBER Working Papers 21473, National Bureau of Economic.* Research, Inc.

Doeringer, P. B. (1993). Who Benefits from State and Local Economic Development Policies? (T. J. Bartik, Ed.) *Price: NPA. J. Pol. Anal. Manage, 12*, 411–413. doi:doi:10.2307/3325252

England, P. (1992). Comparable worth : theories and evidence / Paula England Aldine de Gruyter. New York.

England, P., Allison, P., & Wu, Y. (2007). Does bad pay cause occupations to feminize, Does feminization reduce pay, and How can we tell with longitudinal data? *Social Science Research, 36*, 1237–1256. doi:DOI: 10.1016/j.ssresearch.2006.08.003

Flory, J., Leibbrandt, A., List, & John. (2015). Do Competitive Workplaces Deter Female Workers? *A Large-Scale Natural Field Experiment on Job Entry Decisions, Review of Economic Studies, 82*, 122–155.

Georgellis, Y. W., & H. (2005). Gender differences in self-employment. *International Review of Applied Economics, 19*, 321–337.

Giorgi, G., Paccagnella, M., & Pellizzari, M. (2013). "Gender complementarities in the labor market," Questioni di Economia e Finanza (Occasional. In *Papers) 183, Bank of Italy, Economic Research and International Relations Area.*

Glenn Loury, C. S. (1993). Will Affirmative-Action Policies Eliminate Negative Stereotypes? *The American Economic Review.*

Goldin, C. (2006). The Quiet Revolution That Transformed Women's Employment, Education, and Family. *The American Economic Review, 96*, 1–21.

Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *American Economic Review, 2014*, 1091–1119.

Goldin, C. (2015). A Pollution Theory of Discrimination: Male and Female Differences in Occupations and Earnings. In *Human Capital in History: The American Record* (pp. 313–348). Chicago, IL: University of Chicago Press.

Goldin, C. D. (1988). National Bureau of Economic Research. Cambridge, Mass: National Bureau of Economic Research.

Goldin, C., & Katz, L. F. (2011). Putting the "Co" in Education: Timing, Reasons, and Consequences of College Coeducation from 1835 to the Present,. *Journal of Human, 5*, 377–417.

Goldin, C., & Katz, L. F. (2011). The Cost of Workplace Flexibility for High-Powered Professionals. *The Annals of the American Academy of Political and Social Science, 2011*, 45–67.

Goldin, C., & Katz, L. F. (n.d.). A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation. *Journal of Labor Economics [Internet], 2016*, 3–705.

Goldin, C., Katz, L. F., & Kuziemko, I. (2006). The Homecoming of American College Women: The Reversal of the College Gender Gap. *The Journal of Economic Perspectives, 20*, 133–56. Retrieved from http://www.jstor.org/stable/30033687.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2018). Bartik Instruments: What, When, Why, and How,. In *NBER Working Papers 24408, National Bureau of Economic*. Research, Inc.

Görlich, D., & De Grip, A. (2007). *Human capital depreciation during family-related career interruptions in male and female occupations, Kiel Working.* Paper, No. 1379, Kiel Institute for the World Economy (IfW), Kiel.

Guryan, J. C., & K., K. (2013). Taste-based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots. *Econ J, 123*, 417–432. doi:doi:10.1111/ecoj.12080

Helen Levy, T. D. (2004, 10). Worker Sorting and the Risk of Death on the Job,. *Journal of Labor Economics, 22*, 925–953.

Jaeger, D. A. (1997). Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers. *Journal of Business & Economic Statistics, 15*, 300–309. Retrieved from www.jstor.org/stable/1392334.

Karlin, C. A., England, P., & Richardson, M. G. (n.d.). *Issues, 2002*, 10–1007.

Larson-Koester, M. (n.d.). Occupation Gender Segregation: Empirical Evidence from a Matching Model with Transfers. *(Job Market Paper)*.

Levanon, E., & Allison. (2009). Occupational Feminization and Pay: Assessing Causal Dynamics Using 1950-2000 U.S. Census Data. *Social Forces*.

Lovenheim, M. F., & Reynolds, C. L. (2011). Changes in Postsecondary Choices by Ability and Income: Evidence from the National Longitudinal Surveys of Youth. *Journal of Human Capital, 5*, 70–109. Retrieved from www.jstor.org/stable/10.1086/660123.

Macpherson, D. A., & Hirsch, B. T. (1995). Wages and Gender Composition: Why Do Women's Jobs Pay Less? *Journal of Labor Economics, 13*, 426–471. Retrieved from www.jstor.org/stable/2535151.

Mas, A., & Pallais, A. (2017). Valuing Alternative Work Arrangements. *American Economic Review,107, 12*, 3722–59.

Maume, D. J. (1999, 6). Occupational Segregation and the Career Mobility of White Men and Women. *Social Forces, 77*, 1433–1459. Retrieved from https://doi.org/10.1093/sf/77.4.1433

McClelland, R., Mok, S., & Pierce, K. (2014). Labor Force Participation Elasticities of Women and Secondary Earners within Married Couples: Working Paper 2014-06,. *Reports, 49433*.

Meyer, P. B., & Osborne, A. M. (2005). *"Proposed Category System for 1960-2000 Census Occupations," Working Papers 383.* U.S. Bureau of Labor Statistics.

Murphy, E., Oesch, & Daniel. (2015, 2). The Feminization of Occupations and Change in Wages: A Panel Analysis of Britain, Germany and Switzerland. *SOEPpaper, 731*. Retrieved from https://ssrn.com/abstract

Nakao, K. T., & J. (1994). Updating Occupational Prestige and Socioeconomic Scores: How the New Measures Measure Up. *Sociological Methodology, 24*, 1–72.

Niederle, M., & Vesterlund, L. (2005). Do Women Shy Away From Competition? Do Men Compete Too Much?,. In *NBER Working Papers 11474, National Bureau of Economic.* Research, Inc.

Pan, J. (2010). Gender Segregation in Occupations: The Role of Tipping and Social Interactions. *Journal of Labor Economics*.

Parker. (2018). Retrieved from http://pewrsr.ch/2G5booM

Pitts, M. M. (2002). *"Why choose women's work if it pays less? A structural model of occupational choice," FRB Atlanta Working.* Paper 2002-30, Federal Reserve Bank of Atlanta.

Ruggles, S., Genadek, K., Goeken, R., Grover, J., & Sobek, M. (2015). *Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database].* Minneapolis: University of Minnesota.

Siegel, P. M. (1971). Prestige in the American Occupational Structure,.

Solana, K. (2016, 1). 18 Published). Women in dentistry see progress, continued challenges. Retrieved from https://www.ada.org/en/publications/ada-news/2016-archive/january/women-in-dentistry-see-progress-continued-challenges

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin, 135*, 859–884.

Sullivan, P. (n.d.). Empirical Evidence on Occupation and Industry Specific Human Capital. *Labour economics, 3*. doi:doi:10.1016/j.labeco.2009.11.003

Tam, T. (1997). Sex Segregation and Occupational Gender Inequality in the United States: Devaluation or Specialized Training? *American Journal of Sociology*.

Chapter 2:

28 U.S. Code § 44. Appointment, tenure, residence and salary of circuit judges (https://www.law.cornell.edu/uscode/text/28/44)

Algan, Yann & Quoc-Anh Do & Nicolò Dalvit & Alexis Le Chapelain & Yves Zenou, 2015. "How Social Networks Shape Our Beliefs: A Natural Experiment among Future French Politicians," Sciences Po publicationsinfo:hdl:2441/78vacv4udu9, Sciences Po.

American Bar Association Required Disclosures (2013). Enrollment and Degrees Awarded 1963 – 2012 Academic Years. (https://www.americanbar.org/content/dam/aba/administrative/legal_education_and_admissions_to_the_bar/statistics/enrollment_degrees_awarded.pdf)

American Bar Association Standing (2017). American Bar Association Standing Committee on the Federal Judiciary: What it is and How it Works. ISBN: 978-1-62722-880-0

Avery, Christopher, Christine Jolls, Richard A. Posner, and Alvin E. Roth (2007). "The New Market for Federal Judicial Law Clerks." University of Chicago Law Review 74, no. 2: 447–486.

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles. 2006. "Empathy or Antipathy? The Impact of Diversity." American Economic Review, 96(5): 1890-1905.

Boyd, C. L., Epstein, L. and Martin, A. D. "Untangling the Causal Effects of Sex on Judging," American Journal of Political Science (54:2), 2010, pp. 389--411.

Chilton, Adam S. and Levy, Marin K., Challenging the Randomness of Panel Assignment in the Federal Courts of Appeals (October 5, 2015). Duke Law School Public Law & Legal Theory Series No. 2015-1; U of Chicago, Public Law Working Paper No. 529. Available at SSRN: https://ssrn.com/abstract=2520980 or http://dx.doi.org/10.2139/ssrn.2520980

Dahl, G., Kotsadam, A. and Rooth, D.-O. "Does Integration Change Gender Attitudes? The Effect of Randomly Assigning Women to Traditionally Male Teams"(24351), Technical report, National Bureau of Economic Research, 2018.

District Judges 28 U.S. Code § 292, 1986

Epstein, L., Martin, A. D., Segal, J. A., & Westerland, C. L. (2007). The judicial common space. *Journal of Law, Economics, and Organization*, *23*(2), 303-325. https://doi.org/10.1093/jleo/ewm024

Fancine Blau and Lawrence Kahn. 2016. The Gender Wage Gap: Extent, Trends, and Explanations.NBER Working Paper No. 21913.

Fang, Hanming and Andrea Moro. 2010. “Theories of Statistical Discrimination and Affirmative Action: A Survey, in Handbook of Social Economics.” edited by Jess Benhabib, Alberto Bisin, and Matthew Jackson. North-Holland.

Federal Judicial Center (2019). “Biographical Directory of Article III Federal Judges” https://www.fjc.gov/history/judges/biographical-directory-article-iii-federal-judges-export

Glynn, A. N. and Sen, M. (2015), Identifying Judicial Empathy: Does Having Daughters Cause Judges to Rule for Women's Issues?. American Journal of Political Science, 59: 37–54. doi:10.1111/ajps.12118

Goldin, C. (2015). A Pollution Theory of Discrimination: Male and Female Differences in Occupations and Earnings. In Human Capital in History: The American Record (pp. 313–348). Chicago, IL: University of Chicago Press.

Golub, Benjamin and Matthew Jackson. 2012. “How Homophily Affects the Speed of Learning and Best-Response Dynamics.” Quarterly Journal of Economics 127: 1287–1338.

Gramlich, J. "Federal judicial picks have become more contentious, and Trump’s are no exception", Pew Research Center, 2018.

Hess, A. G. ""The Collapse of the House that Ruth Built: The Impact of the Feeder System on Female Judges and the Federal Judiciary, 1970-2014," Journal of Gender, Social Policy & the Law (24:1), 2015, pp. 1-42.

History of the Federal Judiciary . http://www.fjc.gov . Web site of the Federal Judicial Center, Washington, DC.

Judicial Yellow Book: Who's Who in Federal and State Courts. New York, N.Y: Leadership Directories, Inc, 2007-2017. Publication Masters.

La Ferrara, Eliana, Justin Burns and Lucia Corno. 2014. “Interaction, prejudice and performance: evidence from randomly assigned peers in South Africa.” Unpublished Manuscript.

Levy, Marin K. (2017). Panel Assignment in the Federal Courts of Appeals, 103 *Cornell Law Review* 65-116

Levy, Marin K. (2019). “Visiting Judges,” California Law Review, Vol. 107, No. 67, 2019; Duke Law School Public Law & Legal Theory Series No. 2019-11. Available at SSRN: https://ssrn.com/abstract=3328571

Levy, Gilat and Razin, Ronny (2016) The coevolution of segregation, polarised beliefs and discrimination: the case of private vs. state education. American Economic Journal: Microeconomics . ISSN 1945-7669 (In Press)

Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet (2019). *Voteview: Congressional Roll-Call Votes Database*. https://voteview.com/

Mark, Noah & R Harris, Daniel. (2012). Roommate's race and the racial composition of white college students' ego networks. Social science research. 41. 331-42. 10.1016/j.ssresearch.2011.11.012.

McCabe, P. G. "A Guide to the Federal Magistrate Judges System", Technical report, Federal Bar Association, 2014.

McMillon, B. J. "US Circuit Court Judges: Profile of Professional Experiences Prior to Appointment", Technical report, Congressional Research Service, 2014.
NALP (2019). “Clerkship Study Law Student Findings” https://www.nalp.org/clrkfind3
Pérez, Israel Saeta (2016) “gender-guesser 0.4.0” https://pypi.org/project/gender-guesser/
Posner, Richard Roth, A. E., Christopher Avery, and Christine Jolls (2001). "The Market for Federal Judicial Law Clerks." University of Chicago Law Review 68, no. 3: 793–902.
Rhinehart, Lynn K. (1994). “Is There Gender Bias in the Judicial Law Clerk Selection Process,” 83 Geo L.J. 575
Stearns, Maxwell L. and Abramowicz, Michael, "Defining Dicta" (2005). *Faculty Scholarship*. 119. https://digitalcommons.law.umaryland.edu/fac_pubs/119
Washington, Ebonya L. 2008. "Female Socialization: How Daughters Affect Their Legislator Fathers." American Economic Review, 98(1): 311-32.

Chapter 3:

Baumrind, D. (1967). Child care practices anteceding three patterns of preschool behavior. *Genetic Psychology Monographs, 75*(1), 43-88.

Berlin, L. J., Ispa, J. M., Fine, M. A., Malone, P. S., Brooks-Gunn, J., Brady-Smith, C., … Bai, Y. (2009). Correlates and consequences of spanking and verbal punishment for low-income white, african american, and mexican american toddlers. *Child development*, *80*(5), 1403–1420. doi:10.1111/j.1467-8624.2009.01341.x

Ceballo, R., & McLoyd, V. C. (2002). Social support and parenting in poor, dangerous neighborhoods. Child Development, 73(4), 1310–1321. doi:10.1111/1467-8624.00473.

Coates, T.-N. (2015). Between the world and me. Melbourne: The Text Publishing Company.

Cobb-Clark, Deborah A. and Salamanca, Nicolas and Zhu, Anna, Parenting Style as an Investment in Human Development (February 8, 2016). Melbourne Institute Working Paper No. 3/16. Available at SSRN: https://ssrn.com/abstract=2729578 or http://dx.doi.org/10.2139/ssrn.2729578

Cuellar, J., Jones, D.J. & Sterrett, E. J Child Fam Stud (2015) 24: 195. https://doi.org/10.1007/s10826-013-9826-y

Cunha F. (2015) Subjective Rationality, Parenting Styles, and Investments in Children. In: Amato P., Booth A., McHale S., Van Hook J. (eds) Families in an Era of Increasing Inequality. National Symposium on Family Issues, vol 5. Springer, Cham

Dixon, Sara Villanueva et al. "The roles of respect for parental authority and parenting practices in parent-child conflict among African American, Latino, and European American families." *Journal of family psychology : JFP : journal of the Division of Family Psychology of the American Psychological Association (Division 43)* vol. 22,1 (2008): 1-10. doi:10.1037/0893-3200.22.1.1

Doepke, M., and Zilibotti, M.. " Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission." *Econometrica*, **85**(5), 2017, 1331– 71.

Dornbusch, S., Ritter, P., Leiderman, P., Roberts, D., & Fraleigh, M. (1987). The Relation of Parenting Style to Adolescent School Performance. *Child Development, 58*(5), 1244-1257. doi:10.2307/1130618

Fauth, R. C., Leventhal, T. and Brooks-Gunn, J. (2007), Welcome to the Neighborhood? Long-Term Impacts of Moving to Low-Poverty Neighborhoods on Poor Children's and Adolescents' Outcomes. Journal of Research on Adolescence, 17: 249-284. doi:10.1111/j.1532-7795.2007.00522.x

Gayles, J. G., Coatsworth, J. D., Pantin, H. M., & Szapocznik, J. (2009). Parenting and neighborhood predictors of youth problem behaviors within Hispanic families: The moderating role of family structure. Hispanic Journal of Behavioral Sciences, 31(3), 277–296. doi:10.1177/0739986309338891.

Gershoff, E.T., & Grogan-Kaylor, A.A. (2016). Spanking and child outcomes: Old controversies and new meta-analyses. *Journal of family psychology : JFP : journal of the Division of Family Psychology of the American Psychological Association, 30 4*, 453-69 .

Gonzales, N. A., Coxe, S., Roosa, M. W., White, R. B., Knight, G. P., Zeiders, K. H., et al. (2011). Economic hardship, neighborhood context, and parenting: Prospective effects on Mexican-American adolescent's mental health. American

Journal of Community Psychology, 47(1–2), 98–113. doi:10.1007/s10464-010-9366-1.

Guidelines for Student Suspension, Los Angeles Unified School District Policy Bulletin BUL-3819 (2007).

Harlem Children's Zone. "The Cradle through College Pipeline: Supporting Children's Development through Evidence-Based Practices." 2009: https://www.bpichicago.org/wp-content/uploads/2014/07/The-Cradle-through-College-Pipeline.pdf

Heckman, James & Cunha, Flavio (2007). "The Technology of Skill Formation," American Economic Review, American Economic Association, vol. 97(2), pages 31-47, May.

Huynh-Nhu Le, Rosario Ceballo, Ruth Chao, Nancy E. Hill, Velma McBride Murry & Ellen E. Pinderhughes (2008) Excavating Culture: Disentangling Ethnic Differences From Contextual Influences in Parenting, Applied Developmental Science, 12:4,163-175, DOI: 10.1080/10888690802387880

Kohen, D. E., Leventhal, T. , Dahinten, V. S. and McIntosh, C. N. (2008), Neighborhood Disadvantage: Pathways of Effects for Young Children. Child Development, 79: 156-169. doi:10.1111/j.1467-8624.2007.01117.x

Martínez, I., & García, J. (2007). Impact of Parenting Styles on Adolescents' Self-Esteem and Internalization of Values in Spain. The Spanish Journal of Psychology, 10(2), 338-348. doi:10.1017/S1138741600006600

Noah, T. (2016). Born a crime: Stories from a South African childhood. (OKS Print.)

Park, C and Lau, A. S. "Socioeconomic Status and Parenting Priorities: Child Independence and Obedience Around the World," Journal of Marriage and Family 78.1 (2016): 43-59.

Pebley, Ann R., and Sastry, Narayan. 2011. The Los Angeles Family and Neighborhood Survey, Wave 1 (2000-2001) and Wave 2 (2006-2008) [machine-readable data file and documentation]. Santa Monica, CA. RAND Corporation.

Peterson, Christine E., Anne R. Pebley, Narayan Sastry, Karen Yuhas, Bonnie Ghosh-Dastidar, Ann C. Haas, Jesse Gregory, and Marianne P. Bitler, The Los Angeles Family and Neighborhood Survey, Wave 2: User's Guide and Codebook. Santa Monica, CA: RAND Corporation, 2012. https://www.rand.org/pubs/working_papers/WR240z20.html.

Pew Research Center, December 17, 2015, “Parenting in America: Outlook, worries, aspirations are strongly linked to financial situation”

Pinderhughes, E. E., Nix, R., Foster, E. M., Jones, D., & The Conduct Problems Prevention Research Group (2007). Parenting in Context: Impact of Neighborhood Poverty, Residential Stability, Public Services, Social Networks, and Danger on Parental Behaviors. Journal of marriage and the family, 63(4), 941–953. doi:10.1111/j.1741-3737.2001.00941.x

PREVENTIVE MEASURES AND PROCEDURES FOR STUDENTS WHO VIOLATE LAWS REGARDING DRUGS, ALCOHOL, AND TOBACCO, Los Angeles Unified School District Policy Bulletin Bulletin No. Z-73 (2002).

S Howard, Kimberly & Brooks-Gunn, Jeanne. (2009). The Role of Home-Visiting Programs in Preventing Child Abuse and Neglect. The Future of children / Center for the Future of Children, the David and Lucile Packard Foundation. 19. 119-46. 10.1353/foc.0.0032.

S. Zhang, S.G. Anderson / Children and Youth Services Review 32 (2010) 889–895

Sege RD, Siegel BS; Council on Child Abuse and Neglect; Committee on Psychosocial Aspects of Child and Family Health. Effective Discipline to Raise Healthy Children. Pediatrics. 2018;142(6):e20183112 - February 01, 2019

Snyder, Howard N., and Sickmund, Melissa. 2006. Juvenile Offenders and Victims: 2006 National Report. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.

Steinberg, L. , Lamborn, S. D., Dornbusch, S. M. and Darling, N. (1992), Impact of Parenting Practices on Adolescent Achievement: Authoritative Parenting, School Involvement, and Encouragement to Succeed. Child Development, 63: 1266-1281. doi:10.1111/j.1467-8624.1992.tb01694.x

Straus, M. A., & Donnelly, D. A. (1994). Beating the devil out of them: Corporal punishment in American families. New York, NY, US: Lexington Books/Macmillan.

Tapia, Michael, 2010. "Untangling race and class effects on juvenile arrests," Journal of Criminal Justice, Elsevier, vol. 38(3), pages 255-265, May.

Tobin, T. (2011). Race Is Not Neutral: A National Investigation of African American and Latino Disproportionality in School Discipline.

Wallace, J. M., Goodkind, S., Wallace, C. M., & Bachman, J. G. (2008). Racial, Ethnic, and Gender Differences in School Discipline among U.S. High School Students: 1991-2005. The Negro educational review, 59(1-2), 47–62.

Zhang, Saijun and Anderson, Steven G. (2010). Low-income single mothers' community violence exposure and aggressive parenting practices. Children and Youth Services Review, Volume 32, Issue 6, Pages 889-895, ISSN 0190-7409, https://doi.org/10.1016/j.childyouth.2010.02.010.

Appendix A: Supplementary Tables for Chapter 1

Table A.1: Cross-Sectional Association between fraction female and wage

	Log Female Wage			
	(1)	(2)	(3)	(4)
Fraction Female	-0.42*** (0.08)	-0.46*** (0.1)	-0.1 (0.07)	-0.15 (0.14)
Time-Varying Controls	X	X		
Weighted		X		X
Sample Size	294	294	321	321
	Log Male Wage			
	(1)	(2)	(3)	(4)
Fraction Female	-0.48*** (0.06)	-0.61*** (0.12)	0.06 (0.07)	-0.04 (0.14)
Time-Varying Controls	X	X		
Weighted		X		X
Sample Size	294	294	334	334

*Note: Each column reports the estimated effect of fraction female on log wage for workers in the 2010 American Community Survey. Time-varying controls are age and education, as described in the text. Weighted regressions are weighted by the number of total workers in the occupation in 2010. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

Table A.2: Occupations with highest predicted change in percent female from 1960-2010

Occupation	Δ Resid. Instrument	Δ Resid. %Female	% Female (1960)	Δ % Female, 1960-2010	%College 1980	Weighted Average 1980 Gender Ratio
169: Social scientists, n.e.c.	0.150	1%	40%	15%	0.85	-0.07
167: Psychologists	0.133	9%	29%	38%	0.94	0.52
154: Subject instructors (HS/college)	0.129	-5%	23%	22%	0.92	0.28
105: Therapists, n.e.c.	0.123	-2%	59%	19%	0.70	0.85
183: Writers and authors	0.119	16%	17%	41%	0.76	0.27
14: Managers in education and related fields	0.118	23%	22%	42%	0.74	0.12
174: Social workers	0.109	0%	61%	16%	0.69	0.90
194: Art/entertainment performers and related	0.106	4%	21%	25%	0.35	-0.18
166: Economists, market researchers, and survey researchers	0.101	-2%	13%	20%	0.79	-0.49
195: Editors and reporters	0.100	-1%	35%	17%	0.72	0.30
749: Misc textile machine operators	0.097	-34%	78%	-32%	0.02	-0.01
18: Managers of properties and real estate	0.096	0%	29%	23%	0.33	-0.14
436: Cooks, variously defined	0.087	-21%	62%	-14%	0.04	0.78
375: Insurance adjusters,	0.086	41%	13%	62%	0.33	0.90

Occupation	Δ Resid. Instrument	Δ Resid. %Female	% Female (1960)	Δ % Female, 1960-2010	%College 1980	Weighted Average 1980 Gender Ratio
examiners, and investigators						
188: Art makers: painters, sculptors, craft-artists, and print-makers	0.084	-12%	31%	7%	0.45	0.24
745: Shoemaking machine operators	0.083	68%	3%	61%	0.01	1.31
159: Teachers , n.e.c.	0.083	-9%	55%	3%	0.56	0.83
318: Transportation ticket and reservation agents	0.082	31%	20%	50%	0.22	0.38
185: Designers	0.080	19%	15%	37%	0.42	0.18
25: Other financial specialists	0.079	8%	24%	33%	0.48	0.26

Table A.3: Occupations with lowest predicted change in percent female from 1960-2010

Occupation	Δ Resid. Instrument	Δ Resid. %Female	% Female (1960)	Δ % Female, 1960-2010	%College 1980	Weighted Average 1980 Gender Ratio
95: Registered nurses	-0.169	-0.231	97%	-6%	42%	3.25
48: Chemical engineers	-0.166	-0.096	1%	14%	87%	-3.08
57: Mechanical engineers	-0.163	-0.115	0%	8%	66%	-3.91
44: Aerospace engineer	-0.161	-0.126	1%	11%	77%	-3.05
53: Civil engineers	-0.161	-0.092	1%	13%	74%	-3.27
55: Electrical engineer	-0.146	-0.110	1%	10%	70%	-2.80
69: Physicists and astronomers	-0.136	-0.142	2%	16%	87%	-2.27
97: Dietitians and nutritionists	-0.131	-0.191	91%	0%	61%	2.70
59: Not-elsewhere-classified engineers	-0.127	-0.127	0%	12%	71%	-2.93
85: Dentists	-0.125	-0.076	5%	22%	98%	-1.56
207: Licensed practical nurses	-0.103	-0.134	95%	-2%	7%	3.38
56: Industrial engineers	-0.093	-0.062	2%	16%	56%	-1.96
87: Optometrists	-0.080	0.094	5%	37%	96%	-1.72
79: Foresters and conservation scientists	-0.079	-0.021	2%	18%	63%	-2.02
43: Architects	-0.076	-0.004	2%	24%	83%	-1.78
315: Typists	-0.071	-0.238	95%	-5%	7%	3.65
549: Mechanics and repairers, n.e.c.	-0.063	-0.039	1%	1%	3%	-2.98
473: Farmers (owners and tenants)	-0.063	0.023	2%	12%	9%	-1.95

Occupation	Δ Resid. Instrument	Δ Resid. %Female	% Female (1960)	Δ % Female, 1960-2010	%College 1980	Weighted Average 1980 Gender Ratio
853: Excavating and loading machine operators	-0.061	0.012	0%	1%	1%	-4.10
824: Locomotive operators (engineers and firemen)	-0.059	-0.064	0%	5%	6%	-3.74

Table A.4: Correlations between Residual of Primary and Robustness Instruments

	$\widehat{f_{t,J}}$	$\widehat{f_{t,J}}^{1970}$	$\widehat{f_{t,J}}^{M}$
$\widehat{f_{t,J}}$	1.00		
(sample size)	(697)		
$\widehat{f_{t,J}}^{1970}$	0.78	1.00	
(sample size)	(605)	(609)	
$\widehat{f_{t,J}}^{M}$	0.36	0.35	1.00
(sample size)	(697)	(609)	(728)

Table A.5: Main Results With Skill X Year Controls

	Females				Males			
	Log Wage (t)		Log Wage (t+10)		Log Wage (t)		Log Wage (t+10)	
2SLS:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction Female	-0.74**	-0.56*	-1.3***	-1.3***	-0.78***	-0.73***	-0.78***	-0.68***
	(0.37)	(0.38)	(0.39)	(0.39)	(0.26)	(0.27)	(0.25)	(0.25)
Labor Supply	-0.04	-0.01	-0.22**	-0.19**	-0.01	-0.02	-0.03	0.00
	(0.1)	(0.1)	(0.11)	(0.11)	(0.06)	(0.06)	(0.08)	(0.09)
Male Wage			0.21**	0.24**			0.15**	0.14**
			(0.12)	(0.12)			(0.09)	(0.08)
Female Wage			-0.02	-0.02			0.06**	0.06**
			(0.08)	(0.08)			(0.04)	(0.03)
Skill X Year Controls		X		X		X		X
First-Stage:								
Fraction Female	0.75***	0.77***	0.91***	0.95***	0.75***	0.77***	0.89***	0.94***
	(0.13)	(0.14)	(0.15)	(0.16)	(0.13)	(0.14)	(0.15)	(0.17)
Reduced-Form:								
Fraction Female	-0.56**	-0.43	-1.19***	-1.24***	-0.59***	-0.56***	-0.69***	-0.63***
	(0.26)	(0.28)	(0.28)	(0.28)	(0.17)	(0.19)	(0.19)	(0.21)
Sample Size	1575	1575	1276	1276	1575	1575	1283	1283

*Note: Each column reports results from an estimate of equations (14) and (15) in the paper, with Log Mean Wage for workers over the age of 45 as the dependent variable, estimated for males and females. The unit of observation is the occupation X year. Estimates control for age, education and 1980 education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

Table A.6: Main Results Controlling for Future Value of Instrument

	Log Female Wage		Log Male Wage	
2SLS:	(1)	(2)	(3)	(4)
Fraction	-0.52	-0.73	-0.66**	-0.45
Female	(0.56)	(0.61)	(0.34)	(0.37)
Instrument	-0.24	-0.25	-0.1	-0.08
(t+10)	(0.59)	(1.12)	(0.39)	(0.75)
Instrument		-0.21		-0.26
(t+20)		(1.59)		(0.85)
Labor	-0.05	0.01	-0.02	0.04
supply	(0.18)	(0.18)	(0.08)	(0.09)
First-Stage:				
Fraction	1.3***	1.41***	1.23***	1.26***
Female	(0.34)	(0.36)	(0.35)	(0.38)
Reduced-Form:				
Fraction	-0.68	-1.03	-0.81**	-0.56
Female	(0.73)	(0.9)	(0.37)	(0.44)
Sample				
Size	1456	1096	1466	1116

*Note: Each column reports results from an estimate of equations (5) (6) and (7) in the paper, with Log Mean Wage for workers over the age of 45 as the dependent variable, estimated for males and females. I add an estimate of the instrument at t+10 and t+20. The unit of observation is the occupation X year. Estimates control for age, education and 1980 education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

Table A.7: Primary Results on Log Mean Wage of College Graduates

	Females				Males			
	Log Wage (t)	Log Wage (t+10)			Log Wage (t)	Log Wage (t+10)		
2SLS:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction Female	-0.57	-1.27***	-1.2**	-0.9**	-0.95***	-0.9**	-0.71*	-0.63*
	(0.47)	(0.48)	(0.48)	(0.45)	(0.36)	(0.36)	(0.37)	(0.35)
Labor Supply	-0.29**	-0.47***	-0.44***		-0.23**	-0.21**	-0.17*	
	(0.11)	(0.13)	(0.13)		(0.09)	(0.1)	(0.1)	
Female Labor Supply				-0.56***				-0.21*
				(0.14)				(0.11)
Male Labor Supply				-0.19				-0.11
				(0.16)				(0.13)
Log Wage (t)		-0.14***				-0.15***		
		(0.03)				(0.03)		
Controls	X	X	X	X	X	X	X	X
Occ Fe	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X
First-Stage:								
Fraction Female	0.74***	0.77***	0.77***	0.84***	0.74***	0.78***	0.77***	0.83***
	(0.07)	(0.09)	(0.09)	(0.09)	(0.07)	(0.09)	(0.08)	(0.09)
Reduced-Form:								
Fraction Female	-0.42	-0.98***	-0.92***	-0.75**	-0.7***	-0.7***	-0.54**	-0.52*
	(0.34)	(0.34)	(0.35)	(0.36)	(0.25)	(0.26)	(0.27)	(0.28)
Sample Size	1495	1155	1155	1155	1495	1155	1155	1155

*Note: Each column reports results from an estimate of equations (14) and (15) in the paper, with Log Mean Wage for workers over the age of 45 with a bachelor's degree or above as the dependent variable, estimated for males and females. The unit of observation is the occupation X year. Estimates control for age, education and 1980 education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

Table A.8: Primary Results on Workers Age 22-35, Not Enrolled in School

	Females				Males			
	Log Wage (t)	Log Wage (t+10)			Log Wage (t)	Log Wage (t+10)		
2SLS:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction Female	-0.81***	-1.12***	-1.33***	-2.15***	-0.57**	-0.29*	-0.5**	-0.95***
	(0.33)	(0.38)	(0.4)	(0.68)	(0.29)	(0.22)	(0.25)	(0.37)
Labor Supply	-0.3***	-0.32***	-0.33***		-0.17***	-0.16***	-0.18***	
	(0.08)	(0.1)	(0.1)		(0.05)	(0.05)	(0.05)	
Female Labor Supply				-0.63**				-0.4***
				(0.27)				(0.15)
Male Labor Supply				0.24				0.08
				(0.23)				(0.12)
Log Male Wage (t)		0.31***				0.24***		
		(0.08)				(0.05)		
Log Female Wage (t)		-0.03				0.04**		
		(0.06)				(0.02)		
Controls	X	X	X	X	X	X	X	X
Occ Fe	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X
First-Stage:								
Fraction Female	0.8***	0.87***	0.87***	0.58***	0.8***	0.85***	0.85***	0.57***
	(0.12)	(0.15)	(0.14)	(0.13)	(0.12)	(0.15)	(0.15)	(0.13)
Reduced-Form:								
Fraction Female	-0.65***	-0.97***	-1.16***	-1.26***	-0.45**	-0.25	-0.42**	-0.54***
	(0.23)	(0.26)	(0.26)	(0.23)	(0.22)	(0.18)	(0.2)	(0.16)
Sample Size	1816	1456	1456	1456	1816	1466	1466	1466

*Note: Each column reports results from an estimate of equations (14) and (15) in the paper, with Log Mean Wage for workers who are not in school and are age 22-35 as the dependent variable, estimated for males and females. The unit of observation is the occupation X year. Estimates control for age, education and 1980 education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

Table A.9: Results without labor supply control, without base-year education composition control

	Females						Males					
	Log Wage (t)			Log Wage (t+10)			Log Wage (t)			Log Wage (t+10)		
2SLS:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Fraction Female	-0.61**	-0.95*	-0.39	-1.14***	-1.86***	-1.01**	-0.77***	-0.92***	-0.41*	-0.81***	-1.02***	-0.43*
	(0.36)	(0.62)	(0.47)	(0.39)	(0.73)	(0.53)	(0.26)	(0.34)	(0.28)	(0.27)	(0.37)	(0.3)
Labor Supply	-0.11			-0.25**			-0.05			-0.07		
	(0.11)			(0.11)			(0.05)			(0.06)		
Log Male Wage (t)				0.23**	0.21**	0.23***				0.13**	0.13*	0.13**
				(0.1)	(0.13)	(0.1)				(0.08)	(0.08)	(0.07)
Log Female Wage (t)				-0.01	-0.02	0.02				0.06**	0.06*	0.08***
				(0.07)	(0.08)	(0.07)				(0.04)	(0.04)	(0.04)
Base-year Education	X	X		X	X		X	X		X	X	
Time-Varying Controls	X	X	X	X	X	X	X	X	X	X	X	X
Occ Fe	X	X	X	X	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X	X	X	X	X
First-Stage:												
Fraction Female	0.8***	0.51***	0.54***	0.87***	0.56***	0.57***	0.8***	0.51***	0.54***	0.85***	0.55***	0.56***
	(0.12)	(0.11)	(0.1)	(0.15)	(0.13)	(0.12)	(0.12)	(0.11)	(0.1)	(0.15)	(0.14)	(0.12)
Reduced-Form:												
Fraction Female	-0.49*	-0.48*	-0.21	-0.99***	-1.04***	-0.58**	-0.62***	-0.47***	-0.22	-0.68***	-0.56***	-0.24
	(0.28)	(0.28)	(0.24)	(0.28)	(0.27)	(0.25)	(0.19)	(0.15)	(0.15)	(0.2)	(0.15)	(0.15)
Sample Size	1816	1816	1816	1456	1456	1456	1816	1816	1816	1466	1466	1466

*Note: Each column reports results from an estimate of equations (14) and (15) in the paper, with Log Mean Wage for workers over the age of 45 as the dependent variable, estimated for males and females. The unit of observation is the occupation X year. Base-Year Controls are the share of the workforce in each education category in 1980, interacted with year. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

Table A.10: Main Results Estimated for Aggregated Occupations

	Females				Males			
	Log Wage (t)		Log Wage (t+10)		Log Wage (t)		Log Wage (t+10)	
2SLS:	(1)		(2)		(4)		(5)	
	-0.80*		-1.39***				-0.52*	
Fraction Female	(0.50)		(0.53)		-0.58	(0.47)	(0.35)	
					0.00		-0.15*	
Labor Supply	-0.14	(0.14)	-0.34**	(0.16)	(0.12)		(0.09)	
Controls	X		X		X		X	
Occ Fe	X		X		X		X	
Year FE	X		X		X		X	
First-Stage:								
	0.8***		0.91***		0.8***		0.94***	
Fraction Female	(0.27)		(0.26)		(0.27)		(0.26)	
Reduced-Form:								
	-0.64**		-1.27***				-0.49	
Fraction Female	(0.31)		(0.35)		-0.46	(0.33)	(0.30)	
Sample Size	458		379		458		381	

*Note: Each column reports results from an estimate of equations (14) and (15) in the paper, with Log Mean Wage for workers over the age of 45 as the dependent variable, estimated for males and females. The unit of observation is Aggregated occupation X year. Estimates control for age, education and 1980 education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

Table A.11: Main Results Estimated with Controls for Employment

	Females				Males			
	Log Wage (t)	Log Wage (t+10)			Log Wage (t)	Log Wage (t+10)		
2SLS:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction Female	-0.64**	-1.13***	-1.3***	-1.90***	-0.72***	-0.73***	-0.87***	-1.09***
	(0.38)	(0.41)	(0.41)	(0.65)	(0.24)	(0.25)	(0.27)	(0.34)
Labor Supply	-0.08	-0.25**	-0.26**		-0.06	-0.09*	-0.10*	
	(0.10)	(0.11)	(0.11)		(0.06)	(0.06)	(0.06)	
Log Employment (t)	0.01				-0.02*			
	(0.02)				(0.01)			
Log Employment (t+10)		0.00	0.00	0.01		-0.03***	-0.04***	-0.05***
		(0.02)	(0.02)	(0.02)		(0.01)	(0.02)	(0.02)
Male Wage		0.23**				0.13**		
		(0.1)				(0.07)		
Female Wage		-0.01				0.06**		
		(0.07)				(0.04)		
First-Stage:								
Fraction Female	0.75***	0.85***	0.85***	0.59***	0.86***	0.9***	0.91***	0.58***
	(0.12)	(0.15)	(0.14)	(0.13)	(0.12)	(0.14)	(0.14)	(0.12)
Reduced-Form:								
Fraction Female	-0.47*	-0.96***	-1.11***	-1.12***	-0.62***	-0.66***	-0.79***	-0.63***
	(0.27)	(0.29)	(0.28)	(0.27)	(0.19)	(0.2)	(0.21)	(0.16)
Sample Size	1816	1456	1456	1456	1816	1466	1466	1466

*Note: Each column reports results from an estimate of equations (14) and (15) in the paper, with Log Mean Wage for workers over the age of 45 as the dependent variable, estimated for males and females. The unit of observation is the occupation X year. Estimates control for age, education and 1980 education of workers, as described in the text. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

Table A.12: 2SLS Estimate of the Long-Term Effect of Gender Composition on Wage

First-Stage:	Δ percent female	Δ percent female
Δ Instrumented percent female	0.60**	0.60**
	(0.24)	(0.24)
Δ Induced LSF	-0.10 *	-0.10 *
	(0.06)	(0.06)
Reduced-Form:	Δ Female Wage	Δ Female Wage
Δ Instrumented percent female	-0.55	-0.21
	(0.46)	(0.27)
Δ Induced LSF	0.17	0.09
	(0.14)	(0.08)
2SLS:	Δ Female Wage	Δ Male Wage
Δ Instrumented percent female	**-0.91**	**-0.36**
	(0.87)	**(0.45)**
Δ Induced LSF	0.08	0.05
	(0.13)	(0.07)
Sample Size:	160	160

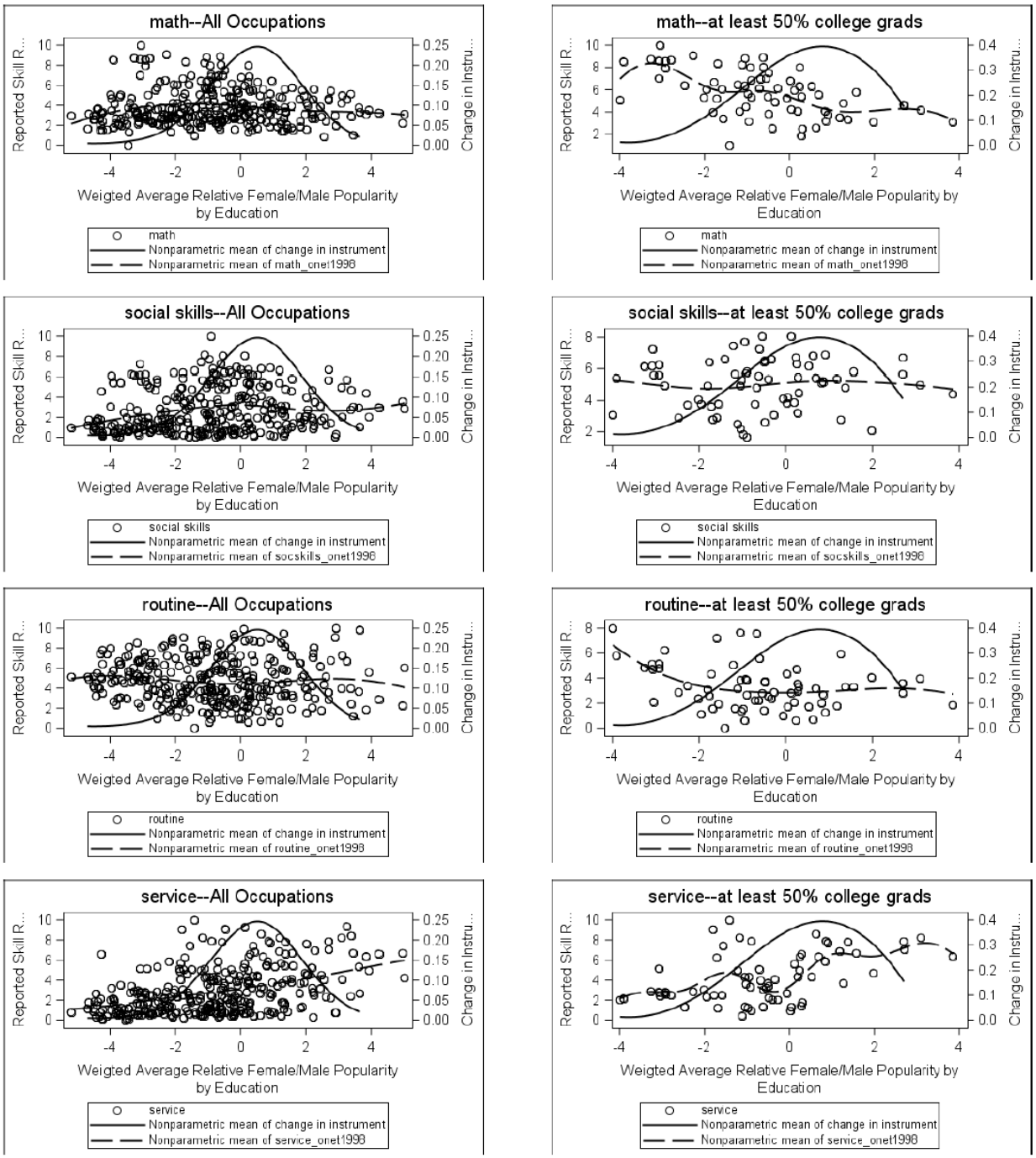
*Note: Each column reports results from an estimate of equations (16) and (17) in the paper, with the change from 1960-2010 in Log Mean Wage for workers over the age of 45 as the dependent variable, estimated for males and females. The unit of observation is the occupation. Controls are described in the paper, and are measured as the difference between 1960 and 2010 values. Standard errors are in parenthesis and are clustered at the occupation level. *** p<0.01, ** p<0.05, * p<0.10

Table A.13: Majors in the American Community Survey

major	Disaggregated majors included	#with major (%with major) among bachelors only	#with major (%with major) among post-bachelors degrees
1: Math/Computer Science	20: Communication Technologies, 21: Computer and Information Sciences, 37: Mathematics and Statistics	70593 (5.3%)	36584 (4.9%)
2: Humanities/ Communications/ Law	19: Communications, 26: Linguistics and Foreign Languages, 32: Law, 33: English Language, Literature, and Composition, 34: Liberal Arts and Humanities, 35: Library Science, 48: Philosophy and Religious Studies, 49: Theology and Religious Vocations, 60: Fine Arts	236101 (17.9%)	103634 (14%)
3: Biological/ Life Sciences	11: Agriculture, 13: Environment and Natural Resources, 36: Biology and Life Sciences	72141 (5.5%)	75984 (10.3%)
4: Medical and Health Services	61: Medical and Health Sciences and Services	91937 (7%)	47942 (6.5%)
5: Social Work/ Interdisciplinary/ Psychology	15: Area, Ethnic, and Civilization Studies, 22: Cosmetology Services and Culinary Arts, 29: Family and Consumer Sciences, 40: Interdisciplinary and Multi-Disciplinary Studies (General), 41: Physical Fitness, Parks, Recreation, and Leisure, 52: Psychology, 54: Public Affairs, Policy, and Social Work	122429 (9.3%)	88923 (12%)

major	Disaggregated majors included	#with major (%with major) among bachelors only	#with major (%with major) among post-bachelors degrees
6: Physical Sciences/ Engineering	14: Architecture, 24: Engineering, 25: Engineering Technologies, 50: Physical Sciences, 51: Nuclear, Industrial Radiology, and Biological Technologies	135441 (10.3%)	102661 (13.9%)
7: Construction/ Manufacturing/ Criminal Justice	38: Military Technologies, 53: Criminal Justice and Fire Protection, 56: Construction Services, 57: Electrical and Mechanic Repairs and Technologies, 58: Precision Production and Industrial Arts, 59: Transportation Sciences and Technologies	43179 (3.3%)	9792 (1.3%)
8: Education	23: Education Administration and Teaching	111569 (8.4%)	90436 (12.2%)
9: Social Science	55: Social Sciences, 64: History	114824 (8.7%)	86304 (11.7%)
10: Business	62: Business	322958 (24.4%)	97929 (13.2%)

Figure A.1: Relationship Between Weighted Female/Male Popularity Ratio and Skills



Appendix B:
Chapter 1 Data Appendix

Decennial Census and American Community Survey (1960-2010):

I use the decennial census to measure the mean wage and demographic composition of each occupation in the years from 1960-2000, and the American Community Survey in 2010. I define key variables in the following way:

Occupation: I define occupation using a modification of the 1990 occupation codes constructed by IPUMS. These codes are constructed using crosswalks to earlier and later occupation coding schemes, as described by Myer and Osborne (2005). Each crosswalk is constructed by assigning a group of workers from a census year both to an occupation in the previous scheme and an occupation in the new scheme. The new occupation codes are then linked to the old occupation codes by choosing the old code most frequently selected by those with each new code. Because the Standard Occupation Codes underwent a significant reclassification in 1970 and in 2000, the 1990 occupation scheme was chosen to avoid transforming any year's data with two major reclassifications.

I deviate from the 1990 harmonized codes in the 2000 and 2010 census by using the gender-specific occupation classifications provided by Blau, Brummund and Liu (2013). Blau et. al. observed that because the most prevalent 1990 codes for each 2000 code differ for male and female workers, gender segregation of occupations appears to decrease in 2000 when measured with harmonized 1990 occupation codes. To address this, Blau et. al. define a gender-specific crosswalk between the 1990 and 2000 codes

by choosing the most prevalent 1990 occupation separately for women and men for each 2000 occupation code.

Wage: I calculate wage in the decennial census by dividing wage and salary income by estimated annual hours worked. The census asks respondents the number of hours they worked in the past week and the number of weeks they worked in the last year. Respondents choose weeks worked per year as intervals. I use the midpoint of each interval as the respondent's weeks worked. The census changed its question on hours worked after the 1990 census. From 1940-1990, the census asked respondents how many hours they worked last week, and provided their responses as intervals: 1-14, 15-29, 30-34, 35-39, 40, 41-48, 49-59, 60+. From 1980-2010, they asked respondents how many hours per week they usually work, with allowed responses ranging from 0-99. I estimated hours worked per year from 1960-1990 by assigning each individual to the midpoint of each interval, and multiplying hours per week by weeks per year. I estimated hours worked per year in 2000 and 2010 by multiplying usual hours worked per week by weeks worked per year.

Education: I divide respondents into five categories by educational attainment, using the detailed educational attainment variable. For years prior to 1990, the categories were defined as follows:

1) Less than high school: Respondents with no schooling completed, less than 12 grade completed, or 12 grade completed with no high school diploma.
2) High school graduate: High school graduate or GED recipient, less than 2 years of college and no college degree.
3) Some College: 2-3 years of college
4) College graduate: 4-5 years of college
5) 6+ years of college

From 1990 onward, they were defined as:

1) Less than high school: Respondents with no schooling completed, less than 12 grade completed, or 12 grade completed with no high school diploma.
2) High school graduate: High school graduate or GED recipient, less than 2 years of college and no college degree.
3) Some College: 2-4 years of college, no bachelor's degree
4) College graduate: 4+ years of college, no master's, graduate or professional degree
5) Master's, graduate or professional degree

These category definitions follow Jaeger (1997), who finds that these definitions minimize the distance between 1980 and 1990 education definitions. While there are no discontinuities in the gender composition of the some college, college graduate, and master's degree categories from 1980 to 1990, the percent of the workforce in each of these categories does change, with fewer advanced degree holders identified in the 1990 census than in the 1980 census. This is of potential importance to the construction of the labor supply control discussed in Section 1.3.1, but does not appear to affect the residual of labor supply once controlling for year fixed effects.

American Community Survey (2009-2014)

Beginning in 2009, the American Community Survey asked respondents with a college degree to provide their college major, coded into 38 categories. Respondents were allowed to list up to three majors. Because the data demands of calculating the fraction of workers in each occupation with a particular education type are high, I collapsed the original 38 categories into ten broader categories, listed in Appendix Table 1.1.

Occupation was determined from the American Community Survey using the Blau modification of the 1990 occupation codes described for the decennial census.

Panel Study of Income Dynamics:

The Panel Study of Income Dynamics (PSID) surveys a nationally representative sample of 18,000 respondents selected in 1968 and their descendants. Surveys are annual from 1968-1997, and semi-annual from 1999-2015. I use the Panel Study of Income Dynamics to look at the effects of changing gender composition on entry to and exit from occupations. The primary variables used in the PSID are occupation and wage income.

Occupation: The PSID uses 1970 occupation codes from 1968 to 1999, after which it switches to 2000 occupation codes. I crosswalk these codes to the modified 1990 occupation codes described above, and then aggregated based on aggregated occupations listed in the 1990 SOC codebook. A worker is determined to have changed occupation when their occupation differs from their occupation in the previous year.

Wage and Salary Income: Analyses on the PSID are conducted on Wage and Salary Income, rather than calculated hourly wage. This was done because the definition of wage and salary income is more consistent across years of the PSID than are definitions of hours and weeks worked, or direct questions on hourly pay. Wage and Salary income includes the labor component of business and farm income, wage income, bonuses, overtime and commissions, and income from professional practice.

Appendix C: Supplementary Tables for Chapter 2

Appendix Table C.1: Validation of Case and Judge Records Data

court	Published Cases in Administrative Records (1)	Case Records in Database (2)	% of Records Included in Database (3)	Number of Judges: Judicial Yellow Books (4)	Number of Judges: Federal Judicial Center (5)
First Circuit	350	323	92.54	9.6	10.4
Second Circuit	310	301	95.25	22.4	24.3
Third Circuit	262	153	69.90	21.5	22.6
Fourth Circuit	205	192	94.17	13.9	17.4
Fifth Circuit	430	402	94.74	22.1	22.6
Sixth Circuit	364	350	95.94	28.1	29.1
Seventh Circuit	623	592	94.96	14.6	15.5
Eigth Circuit	608	575	94.11	20.4	20.1
Ninth Circuit	638	590	91.86	48.3	47.3
Tenth Circuit	309	274	89.11	19.8	21.4
Eleventh Circuit	280	271	97.38	16.0	17.1
District of Columbia Circuit	220	209	95.32	15.5	14.9
Federal Circuit	-	-	-	15.9	17.0
Total	4600	4232	91.99	268.0	279.6

Notes: Column (1) reports the total number of published cases terminated in each appellate circuit from 2007 to 2017, according to US Judicial Business Statistics. Column (2) reports the number of cases taken from the leagle database from January 1, 2007 to December 31, 2017. Column (3) gives the percent of cases reported in administrative data that are included in the leagle database

Appendix Table C.2: Sample Selection

Sample	Observations	Appellate Judges
Appellate Judges in Judicial Yellow Books, 2007-2017	2076	506
Appearing on at least one published appellate panel, 2007-2017	2164	298
Hired at least one clerk in year after appearing on panel	1074	215
With Known Gender	1074	215
With known race and age	1074	215
Appointed by President (Non-Magistrate Judge)	1074	215
With known staff % female	987	197

Notes: This table reports the number of observations (judge X year) and the number of distinct appellate judges included under the analysis under each set of restrictions imposed on the data. The primary sample consists of 1074 observations from 215 judges. Source: Judicial yellow books, case dataset collected by authors (see data section).

Appendix Table C.3: Characteristics of Judges and Clerks in Sample

	Full Sample Mean (SD) (1)	Male Mean (SD) (2)	Female Mean (SD) (3)	Full Sample Min / Max (4)	Sample Size Male / female (5)	M/F Diff (6)
Judge Characteristics						
Female	0.2598 (0.4387)	0.0000 (0.0000)	1.0000 (0.0000)	0 / 1	795 / 279	***
Hispanic	0.0447 (0.2067)	0.0553 (0.2288)	0.0143 (0.1191)	0 / 1	795 / 279	***
Age (Decades)	6.3389 (1.0125)	6.4479 (1.0304)	6.0283 (0.8910)	3.6 / 9.2	795 / 279	***
Decades on Current Court	1.5996 (1.0173)	1.7119 (1.0153)	1.2796 (0.9546)	-0.7 / 4.4	795 / 279	***
Ideology Score	0.0635 (0.3612)	0.0869 (0.3540)	-0.0031 (0.3739)	-0.521 / 0.693	795 / 279	***
Republican	0.5279 (0.4995)	0.5660 (0.4959)	0.4194 (0.4943)	0 / 1	795 / 279	***
Number of Clerks Hired in Year	2.8818 (1.2181)	2.8239 (1.2031)	3.0466 (1.2471)	1 / 7	795 / 279	***
Number of Cases Heard in Year	56.2793 (39.4288)	56.4881 (39.9463)	55.6846 (37.9794)	1 / 194	795 / 279	
% of years with at least one female clerk hired(a)	0.7030 (0.4572)	0.6906 (0.4626)	0.7384 (0.4403)	0 / 1	795 / 279	
% Of Clerks Female	0.4190 (0.4935)	0.4153 (0.4929)	0.4290 (0.4952)	0.0000 / 1.0000	2870 / 1035	

Notes: This table presents the average values of judge-level covariates in the analysis sample. For all variables other than % of clerks female, the sample is at the judge by year level. For % of clerks female, the sample is at the clerk by year level. a: of years where any clerk is hired Source: Judicial yellow books, case dataset collected by authors (see data section).

Appendix Table C.4: Raw and residual variation in female hires, female co-panelists

Panel A: fraction of co-panelists who are female	Mean	Std.dev.	Min	Max	Obs
Raw variable	0.23	0.11	0.00	1.00	1074
Resuduals: net of court by year fixed effects	0.00	0.08	-0.24	0.77	1074
Residuals: net of court by year fixed effects, judge characteristics	0.00	0.08	-0.25	0.75	1074
Residuals: net of court by year fixed effects, judge fixed effects	0.00	0.06	-0.23	0.29	1074

Panel B: Hired a female clerk	Mean	Std.dev.	Min	Max	Obs
Raw variable	0.70	0.46	0.00	1.00	1074
Resuduals: net of court by year fixed effects	0.00	0.43	-0.92	0.86	1074
Residuals: net of court by year fixed effects, judge characteristics	0.00	0.42	-0.92	0.80	1074
Residuals: net of court by year fixed effects, judge fixed effects	0.00	0.36	-0.96	0.86	1074

Notes: This table reports descriptive statistics for the key dependent and independent variables in this analysis. The key dependent variable is an indicator for whether a judge hired a female clerk in each year, conditional on hiring. The key independent variable is the fraction of a judge's co-panelists who were rated as highly qualified by a majority of American Bar Association Raters who are female in each year. Judge characteristics include quadratics of judge age, experience in current position and ideology, judge gender, hispanic ethnicity and party of nominating president. Source: Judicial yellow books, case dataset collected by authors (see data section).

Appendix Table C.5: Escalating interaction controls

Dep Var: Probability of hiring any female clerk in next year	(1)	(2)	(3)	(4)	(5)	(6)
Fraction of co-panelists who are female	0.3972**	0.3881**	0.3854**	0.3506*	0.4092**	0.3464*
	(0.1845)	(0.1843)	(0.1902)	(0.1873)	(0.1868)	(0.1921)
Fraction of co-panelists who are Republican		-0.0571				-0.1107
		(0.1422)				(0.1550)
Fraction of co-panelists <10 years in current position			0.2564			0.1942
			(0.1619)			(0.2048)
Fraction of co-panelists <60 years old				0.2615*		0.1500
				(0.1534)		(0.1855)
Fraction of co-panelists above average citations					0.0562	0.0121
					(0.1438)	(0.1491)
Judge Controls	No	Yes	Yes	Yes	Yes	Yes
Court X Year X District Judge FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1074	1074	1074	1074	1074	1074
Dependent Variable Mean	0.7030	0.7030	0.7030	0.7030	0.7030	0.7030

Notes: This table reports estimated coefficients from the regressions described in equations (1) and (2) in the text. The dependent variable is an indicator of whether a judge hired at least one female clerk in the following year, conditional on hiring any clerk. The table reports the regression of the dependent variable on the fraction of co-panelists who were female in each year. Column (5) adds controls for the fraction of co-panelists who are republican, who are younger than 60 years old, who have served fewer then 10 years on the court, who have a current staff that is more than 50% female, and who have an above average citation rate. Significance levels are: * 10%, ** 5%, *** 1%. Source: Judicial yellow books, case dataset collected by authors (see data section).

Appendix D: Supplementary Tables for Chapter 3

Appendix Table D.1: Definitions of Parenting and Adult Measures

Panel A: Actual discipline	Wave 1		Wave 2	
	Weight	Corr	Weight	Corr
	(1)	(2)	(3)	(4)
# Times spanked	0.23	0.44	0.22	0.34
# Times grounded	0.40	0.77	0.39	0.82
# Times took away privileges	0.41	0.82	0.39	0.84
# Times sent to room	0.38	0.78	0.36	0.76
Panel B: Intended Harsh discipline	(1)	(2)	(3)	(4)
Would spank child if had a temper tantrum	0.60	0.68	0.68	0.73
Would send child to room for more than 30 minutes if had a temper tantrum	0.40	0.39	0.68	0.74
Would scold child if had a temper tantrum	0.62	0.71		
Panel C: Parent: Friends in Neighborhood	(1)	(2)	(3)	(4)
# Adults you recognize in neighborhood (rev)	-0.22	-0.53	-0.18	-0.52
# Neighbors talked to in 30 days (rev)	-0.28	-0.74	-0.21	-0.65
How close do you feel toward neighbors	0.27	0.72	0.24	0.77
How often neighbors do favors for each other?	0.27	0.72	0.22	0.71
How often do neighbors ask you for advice?	0.25	0.68	0.22	0.69
# Friends living in neighborhood	-0.22	-0.55	0.20	0.60
Panel D: Parent: Participates in Neighborhood	(1)	(2)	(3)	(4)
Participated in neighborhood meeting	0.21	0.58	0.19	0.55
Participated in business/civic group	0.24	0.59	0.20	0.51
participated in nationality/ethnic pride club	0.21	0.48	0.13	0.33
Participated il local/state political org.	0.25	0.60	0.20	0.54
Volunteered with local org	0.27	0.75	0.25	0.75
Participated in veterans group	0.09	0.18	0.10	0.23
Participated il labor union	0.12	0.28	0.10	0.26
Participated in literary/art discussion group	0.23	0.59	0.21	0.59
Participated in fraternity/sorority	0.19	0.44	0.15	0.39

Appendix Table D.1 (Cont): Definitions of Parenting and Adult Measures

Panel E: Parent: Participates in Neighborhood	(1)	(2)	(3)	(4)
How Satisfied with neighborhood	0.13	0.62	0.13	0.66
Have been robbed in neighborhood	-0.06	-0.32	-0.05	-0.28
Safe to walk around alone	0.12	0.56	0.12	0.62
Close-knit neighborhood	0.13	0.65	0.13	0.66
Adults kids can look up to	0.13	0.64	0.13	0.67
People are willing to help neighbors	0.13	0.65	0.14	0.67
Neighbors generally don’t get along	-0.09	-0.46	-0.09	-0.46
People don't share the same values	-0.07	-0.35	-0.08	-0.39
People in neighborhood can be trusted	0.15	0.72	0.14	0.72
Parents know kids friends	0.12	0.60	0.11	0.52
Adults know local kids	0.11	0.51		
Parents know each other	0.11	0.54		
Neighbors do something if kid hangs out	0.12	0.63	0.12	0.65
Would do something if kid does graffiti	0.13	0.68	0.13	0.68
would scold kid if showing disrespect	0.10	0.51	0.11	0.56
Adults watch out that kids are safe			0.14	0.68

Notes: Table reports the weights used in constructing adult and parenting measures in columns 1 and 3, and the correlation of each construct with the constituent element in columns 2 and 4.

Appendix Table D.2: Definitions of Child and School Measures

Panel A: Youth-Reported Delinquency	Children < 12				Children >= 12			
	Wave 1		Wave 2		Wave 1		Wave 2	
	Weight (1)	Corr (2)	Weight (3)	Corr (4)	Weight (1)	Corr (2)	Weight (3)	Corr (4)
Ever skipped school without permission (rev)			-0.51	-0.99	-0.17	-0.24	-0.14	-0.85
freq skipped school?	1.00	1.00	0.51	0.95	0.18	0.24	0.15	0.86
Ever sold drugs?					0.14	0.14	0.10	0.41
misbehave at school (rev)					-0.09	-0.11	-0.07	-0.37
Have trouble getting along with teachers (rev)					-0.08	-0.12	-0.06	-0.30
Ever smoked a cigarette					0.16	0.28	0.15	0.58
Ever drank acohol					0.18	0.77	0.13	0.56
# Days in past month drinking alcohol					0.12	0.99	0.09	0.52
Ever used marijuana					0.18	0.26	0.15	0.57
Drugs other than marijuana					0.11	0.08	0.14	0.45
Ever run away from home							0.11	0.43
Ever carry a gun (30 days)					0.09	0.04	0.08	0.22
Ever belonged to a gang					0.09	0.10	0.06	0.17
Ever had sexual intercourse					0.15	0.23	0.13	0.45
Ever pregnant/gotten someone pregnant					0.08	0.09	0.09	0.25
Ever been arrested							0.12	0.41
Panel B: Youth-Reported School Quality	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Are teachers good at your school?	0.55	-0.77			0.45	-0.79	0.47	-0.79
Do teachers care about students?	0.54	-0.70	0.67	-0.56	0.47	-0.84	0.48	-0.81
Are students treated fairly?	0.32	-0.58	0.67	-0.90	0.36	-0.70	0.36	-0.68
Panel B: Youth-Reported School Quality	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Do other students misbehave?	-0.67	-0.84	-0.68	-0.83	-0.65	-0.84	-0.65	-0.84
Do you feel safe at this school?	0.67	0.64	-0.68	-0.63	0.65	0.69	0.65	0.66

Notes: Table reports the weights used in constructing child-reported measures in columns 1, 3, 5, and 7, and the correlation of each construct with the constituent element in columns 2, 4, 6, and 8. Columns 1-4 report on construction of measures for young children, and columns 5-8 report on construction of measures for older children.

Appendix Table D.3: Correlation between Parenting Measure

Corr [P-value] Sample Size		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Actual Discipline	(1)	1.000 [0.000] 1158									
Intended Harsh Discipline	(2)	0.121 [0.000] 1154	1.000 [0.000] 1154								
Strict Discipline	(3)	0.156 [0.000] 1154	0.134 [0.000] 1154	1.000 [0.000] 1154							
Consistent Discipline	(4)	0.057 [0.054] 1154	0.166 [0.000] 1154	0.286 [0.000] 1154	1.000 [0.000] 1154						
Passive Discipline	(5)	-0.078 [0.008] 1154	-0.088 [0.003] 1154	-0.087 [0.003] 1154	0.093 [0.002] 1154	1.000 [0.000] 1154					
Parental Investment in Home Environment	(6)	-0.035 [0.308] 857	-0.058 [0.089] 854	-0.039 [0.259] 854	0.044 [0.202] 854	0.123 [0.000] 854	1.000 [0.000] 1552				
Parental Investment in School	(7)	0.048 [0.100] 1155	-0.003 [0.928] 1151	-0.018 [0.547] 1151	0.079 [0.007] 1151	0.035 [0.241] 1151	0.242 [0.000] 1551	1.000 [0.000] 1996			
Youth-Reported Parental Monitoring	(8)	0.080 [0.020] 854	0.031 [0.369] 852	0.046 [0.182] 852	0.004 [0.896] 852	-0.037 [0.279] 852	0.157 [0.000] 1244	0.169 [0.000] 1509	1.000 [0.000] 1512		
Youth-Reported Parental Warmth	(9)	-0.082 [0.017] 857	-0.046 [0.181] 855	-0.030 [0.385] 855	0.011 [0.750] 855	0.048 [0.161] 855	0.125 [0.000] 1246	0.008 [0.753] 1513	-0.213 [0.000] 1512	1.000 [0.000] 1516	
Youth-Reported Admiration for Parent	(10)	-0.033 [0.342] 856	-0.035 [0.301] 854	-0.030 [0.386] 854	0.028 [0.416] 854	0.056 [0.100] 854	0.153 [0.000] 1248	0.019 [0.471] 1512	-0.047 [0.067] 1510	0.375 [0.000] 1512	1.000 [0.000] 1515

Notes: This table reports correlations between parenting constructs. Columns 1-10 represent correlations with parenting measures 1-10

Appendix Table D.4: Correlation Between Child and Parent Measure

Panel A: Correlation Between Child Measures					
Corr					
[P-value]					
Sample Size		(1)	(2)	(3)	(4)
Youth-Reported Delinquency	(1)	1.000			
		[0.000]			
		1516			
Youth-Reported Study Effort	(2)	0.007	1.000		
		[0.776]	[0.000]		
		1508	1523		
Parent-Reported Youth Delinquency	(3)	0.037	-0.041	1.000	
		[0.147]	[0.111]	[0.000]	
		1516	1523	1999	
Parent-Reported Youth Study Effort	(4)	0.040	-0.099	-0.206	1.000
		[0.241]	[0.004]	[0.000]	[0.000]
		857	861	1158	1158
Panel B: Correlation Between Child and Parent Measures					
Actual Discipline		0.073	-0.068	0.423	-0.072
		[0.034]	[0.045]	[0.000]	[0.014]
		857	861	1158	1158
Intended Harsh Discipline		0.004	0.014	0.144	-0.107
		[0.918]	[0.690]	[0.000]	[0.000]
		854	858	1154	1154
Strict Discipline		0.020	-0.001	0.068	-0.071
		[0.564]	[0.976]	[0.021]	[0.016]
		854	858	1154	1154
Consistent Discipline		-0.040	-0.019	0.021	-0.015
		[0.240]	[0.573]	[0.484]	[0.619]
		854	858	1154	1154
Passive Discipline		-0.027	-0.043	-0.173	0.080
		[0.428]	[0.207]	[0.000]	[0.007]
		854	858	1154	1154
Parental Investment in Home Environment		-0.004	-0.077	-0.097	0.203
		[0.875]	[0.007]	[0.000]	[0.000]
		1237	1247	1552	857
Parental Investment in School		-0.022	-0.062	-0.004	0.165
		[0.383]	[0.015]	[0.846]	[0.000]
		1513	1522	1996	1155
Youth-Reported Parental Monitoring		0.118	-0.528	0.034	0.082
		[0.000]	[0.000]	[0.187]	[0.017]
		1498	1505	1512	854
Youth-Reported Parental Warmth		-0.275	0.015	-0.105	0.041
		[0.000]	[0.572]	[0.000]	[0.235]
		1501	1509	1516	857
Youth-Reported Admiration for Parent		-0.125	-0.027	-0.062	0.074
		[0.000]	[0.289]	[0.015]	[0.030]
		1501	1508	1515	856

Notes: This table reports correlations between child and parenting constructs. Columns 1-4 represent correlations with child measures 1-4

Appendix Table D.5: School Safety on Intended Harsh Discipline

Dep Var: Intended Harsh Discipline	(1)	(2)	(3)	(4)	(5)	(6)
Reported School Safety	0.03	0.05	0.05	0.05	0.04	0.06
	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.07)
Child < 8				0.05	0.08	-0.04
				(0.16)	(0.23)	(0.26)
Child Age				0.06*	0.11**	0.05
				(0.04)	(0.05)	(0.06)
Household Earnings				-0.00	0.00	0.00
				(0.00)	(0.00)	(0.00)
Parents Married & Living Together				0.10	-0.03	0.03
				(0.09)	(0.12)	(0.14)
Income relative to poverty				-0.00	0.00	-0.04
				(0.07)	(0.08)	(0.10)
Below poverty line				-0.04	0.04	0.24
				(0.11)	(0.14)	(0.17)
Below 200% of poverty line				-0.13	-0.23	-0.48**
				(0.16)	(0.18)	(0.23)
Prim.Parent HS Grad				-0.15	-0.32**	-0.40**
				(0.13)	(0.15)	(0.18)
Prim. Parent Attended College				0.07	0.22	0.31
				(0.15)	(0.17)	(0.20)
Prim. Parent Graduated College				-0.0375	-0.24	-0.47**
					(0.21)	(0.23)
Parent: Friends in Neighborhood					0.04	0.03
					(0.06)	(0.07)
Parent: Participates in Neighborhood					0.00	0.08
					(0.06)	(0.05)
Parent: Feels safe in Neighborhood					0.07	0.03
					(0.07)	(0.08)
Parent: Drinks per week						-0.05**
						(0.02)
Parent: 4 or more drinks at a time per week						0.05
						(0.05)
Parent: reported depression						-0.13
						(0.20)
Wave by Tract FE	No	Yes	Yes	Yes	Yes	Yes
School Quality Control	No	No	Yes	Yes	Yes	Yes
Observations	1119	1119	1119	943	605	434
Dependent Variable Mean	0.0225	0.0225	0.0225	0.0233	0.0005	-0.0002

Notes: This table reports estimated coefficients from the regression described in equation (21). The dependent variable is an index of harsh responses to hypothetical misbehavior. The table reports a regression of intended harsh discipline on average reported misbehavior at the child's school. Column (2) adds controls for census tract by survey wave. Column (3) adds a control for other students' report of school quality. Significance levels are: * 10%, ** 5%, *** 1%.

Appendix Table D.6: School Safety on Actual Discipline

Dep. Var: Actual Discipline	(1)	(2)	(3)	(4)	(5)	(6)
Reported School Safety	-0.04	-0.01	-0.03	-0.01	0.05	0.06
	(0.03)	(0.04)	(0.04)	(0.04)	(0.06)	(0.07)
Child < 8				-0.15	-0.04	-0.08
				(0.17)	(0.24)	(0.26)
Child Age				-0.11***	-0.13***	-0.12**
				(0.03)	(0.05)	(0.06)
Household Earnings				0.00*	0.00**	0.00***
				(0.00)	(0.00)	(0.00)
Parents Married & Living Together				0.14*	0.08	0.15
				(0.08)	(0.11)	(0.14)
Income relative to poverty					-0.11**	-0.16**
				-0.0045	(0.05)	(0.07)
Below poverty line				0.14	0.11	0.04
				(0.10)	(0.13)	(0.16)
Below 200% of poverty line				0.05	0.11	0.30*
				(0.12)	(0.16)	(0.18)
Prim.Parent HS Grad				0.22*	0.17	0.06
				(0.12)	(0.16)	(0.18)
Prim. Parent Attended College				-0.06	-0.11	0.15
				(0.14)	(0.19)	(0.25)
Prim. Parent Graduated College				-0.09	0.02	0.17
				(0.12)	(0.17)	(0.20)
Parent: Friends in Neighborhood					-0.0054	-0.13**
						(0.06)
Parent: Participates in Neighborhood					0.04	0.03
					(0.05)	(0.05)
Parent: Feels safe in Neighborhood					0.11*	0.14**
					(0.06)	(0.07)
Parent: Drinks per week						0.00
						(0.02)
Parent: 4 or more drinks at a time per week						0.07
						(0.05)
Parent: reported depression						0.17
						(0.25)
Wave by Tract FE	No	Yes	Yes	Yes	Yes	Yes
School Quality Control	No	No	Yes	Yes	Yes	Yes
Observations	1123	1123	1123	947	609	436
Dependent Variable Mean	-0.0673	-0.0673	-0.0673	-0.0712	-0.0639	-0.0717

Notes: This table reports estimated coefficients from the regression described in equation (21). The dependent variable is a standardized index of discipline types used in the past week. The table reports a regression of actual discipline on average reported misbehavior at the child's school. Column (2) adds controls for census tract by survey wave. Column (3) adds a control for other students' report of school quality. Significance levels are: * 10%, ** 5%, *** 1%.

Appendix Table D.7: School Disorder on Parent-Reported Delinquency

Dep Var: Parent-reported child delinquency	(1)	(2)	(3)	(4)	(5)	(6)
Reported School Misbehavior	-0.01	-0.02	-0.01	-0.01	-0.01	0.00
	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)	(0.05)
Child < 8				-0.17	0.06	0.11
				(0.12)	(0.15)	(0.17)
Child Age				-0.04	0.01	0.02
				(0.03)	(0.03)	(0.04)
Household Earnings				0.00	0.00	0.00
				(0.00)	(0.00)	(0.00)
Parents Married & Living Together				0.01	-0.03	0.06
				(0.06)	(0.08)	(0.10)
Income relative to poverty				-0.05	-0.02	0.03
				(0.05)	(0.06)	(0.07)
Below poverty line				0.03	-0.04	0.07
				(0.07)	(0.10)	(0.12)
Below 200% of poverty line				0.05	0.08	0.19
				(0.09)	(0.12)	(0.14)
Prim.Parent HS Grad				-0.01	-0.13	-0.18
				(0.09)	(0.12)	(0.13)
Prim. Parent Attended College				0.22*	0.27**	0.32**
				(0.11)	(0.14)	(0.15)
Prim. Parent Graduated College				-0.08	0.04	0.16
				(0.10)	(0.13)	(0.15)
Parent: Friends in Neighborhood					0.04	0.04
					(0.04)	(0.05)
Parent: Participates in Neighborhood					0.04	0.04
					(0.04)	(0.05)
Parent: Feels safe in Neighborhood					0.07	0.03
					(0.04)	(0.05)
Parent: Drinks per week						0.01
						(0.01)
Parent: 4 or more drinks at a time per week						0.05
						(0.04)
Parent: reported depression						0.41**
						(0.19)
Wave by Tract FE	No	Yes	Yes	Yes	Yes	Yes
School Quality Control	No	No	Yes	Yes	Yes	Yes
Observations	1935	1935	1933	1635	1023	744
Dependent Variable Mean	-0.0418	-0.0418	-0.0427	-0.0416	-0.0461	-0.0844

Notes: This table reports estimated coefficients from the regression described in equation (21). The dependent variable is an index of parent-reported child misbehavior. The table reports a regression of child misbehavior on other students' report of misbehavior at the child's school. Column (2) adds controls for census tract by survey wave. Column (3) adds a control for other students' report of school quality. Significance levels are: * 10%, ** 5%, *** 1%.