WHO YOU GONNA CALL? GENDER INEQUALITY IN EXTERNAL DEMANDS FOR PARENTAL INVOLVEMENT *

Kristy Buzard[1], Laura K. Gee[2], and Olga Stoddard[3]

[1]Syracuse University
[2]Tufts University
[3]Brigham Young University

May 26, 2023

Abstract

The gender imbalance in time spent on child rearing causes gender inequalities in a wide range of labor market outcomes, human capital accumulation, and economic mobility. We investigate a novel source of this inequality: external demands for parental involvement. We pair a theoretical model with a large-scale field experiment that we conduct with a near-universe of US schools. School principals receive an email from a two-parent household with a general inquiry and are asked to call one of the parents back. Mothers are 1.4 times more likely than fathers to be contacted. We decompose this inequality into discrimination stemming from differential beliefs about parents' availability versus other factors, including gender norms. Our findings underscore a process through which agents outside the household contribute to within-household gender inequalities.

JEL Classification: J16, J71, C93, J22
Keywords: Discrimination, Gender Gap, Field Experiment

---

*Corresponding Author: Laura K. Gee, 177 College Avenue Medford MA 02155, 617-627-5947, laura.k.gee@gmail.com.
We thank Katelyn Cranney, Mason Fields, Katherine Jolley, Boston Kelley, Paris Rich, and Kiersten Robertson for their excellent research assistance. We are grateful to Amanda Agan, Nageeb Ali, Joshua Angrist, David Autor, Peter Bergman, Fran Blau, Kasey Buckles, Amanda Chuan, Zoe Cullen, Amy Finkelstein, Ingrid Haegele, Christopher Karpowitz, Louis-Pierre Lepage, Stephen O'Connell, Daniele Paserman, Amanda Pallais, Jessica Preece, Nina Roussille, Frank Schilbach, and Betsey Stevenson for their valuable feedback. Special thanks to Francis Ditraglia for many useful conversations. We express our appreciation to audience members at the SITE-Gender, the Economic Science Association Conference, the Discrimination and Disparities Seminar, Behavioral Experimental Design Initiative, Stockholm University (SOFI), Uppsala University, Gothenburg University, Middlebury College, MIT, Montana State University, UMass Amherst, George Mason University, and the Association for Social Sciences (ASSA) meetings. We thank the Brigham Young University College of Family, Home, and Social Sciences, the Syracuse University Maxwell School, and Tufts University for their generous research support. This work has been supported (in part) by Grant 2204-38170 from the Russell Sage Foundation. Any opinions expressed are those of the principal investigators alone and should not be construed as representing the opinions of the funder. This study was pre-registered at the AEA registry (AEARCTR-0007610) and was approved by the relevant IRBs.

1 Introduction

Despite the convergence of men's and women's roles in the labor market, a substantial and persistent gender earnings gap of nearly 18% remains (US Census Bureau, 2020). Many factors contributing to this gap are well-documented in the literature. Of recent focus is women's tendency to concentrate in occupations with more temporal flexibility, which is especially true for women with children (Price and Wasserman, 2022; Duchini and Van Effenterre, 2022; Wasserman, 2022; Goldin, 2014).

The need for greater workplace flexibility is consistent with the robust finding that women—even those who work outside the home—engage in a disproportionate share of child- and household-related tasks.[1] US time-use data reveal that married mothers employed full time spend over 50% more time caring for children and engaging in housework and food preparation than analogous fathers (see panel (a) of Figure 1). Consequently, 35% of mothers report experiencing a household interruption during their workday, compared to only 20% of fathers, costing women 9% in wages (Cubas et al., 2021). These gender imbalances come at a significant economic cost to women, stunting labor market outcomes, human capital accumulation, and economic growth as documented extensively in the motherhood penalty literature.[2]

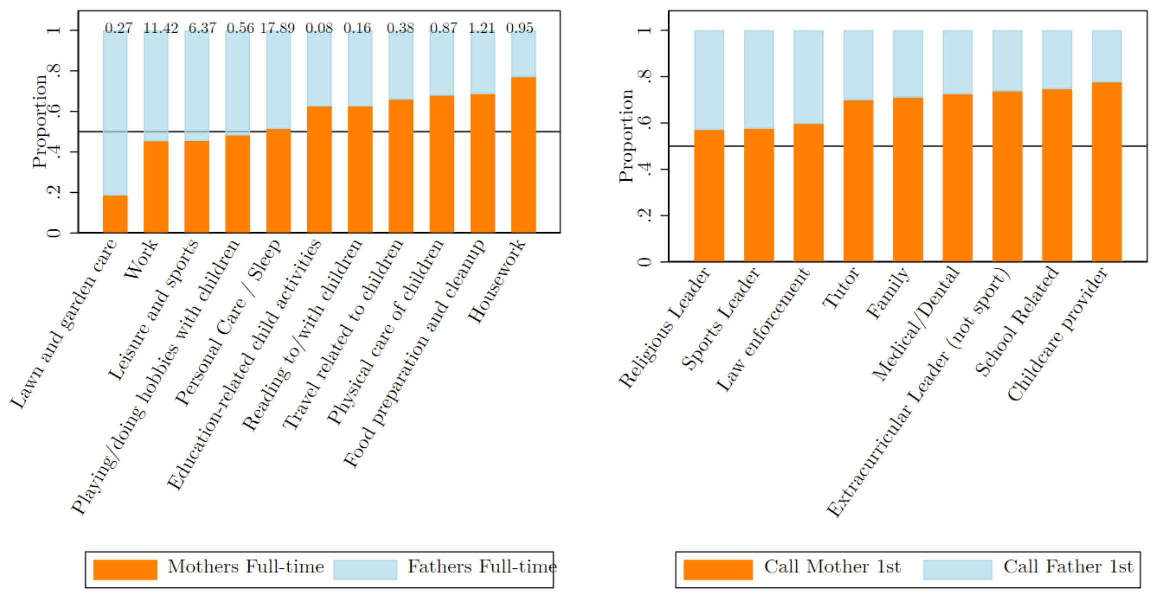
In this paper, we investigate a novel source of this inequality, which we refer to as "external demands for parental involvement." In short, institutions beyond the household and beyond the place of employment put demands on families, and these demands largely fall on mothers. Some of these expectations come from outside forces, such as schools, doctors' offices, church groups, or even grandparents. Small, optimizing decisions by these external agents and countless more create powerful disadvantages for women. Clearly, women anticipate and respond to these external demands by changing the type of work they do and the careers they choose, which ultimately curbs how they progress in those careers. This leads to worse outcomes for society as a whole by reinforcing social biases and perpetuating the cycle of gender inequality. Intuitively, we argue that pervasive societal expectations and social biases curtail women's ability to fully participate in the labor market.

The social biases inherent in the external demands placed on parents can take many forms. Women, for example, get called on more often than men for child-related tasks, such as

[1] See, for example, Aguiar and Hurst (2007); Craig and Mullan (2011); Schoonbroodt (2018).

[2] Many prior studies have documented the motherhood penalty in a wide range of contexts, including work by Adams-Prassl et al. (2023); Ciasullo and Uccioli (2023); Kleven (2023); Speer and Ersoy (2022); Erosa et al. (2022); Albanese et al. (2022); Cubas et al. (2021); Duchini and Van Effenterre (2022); Cubas et al. (2022); Kleven et al. (2019); Kuziemko et al. (2018) and Angelov et al. (2016).

Figure 1: Gender Inequality in Household Time Use and External Contacts



(a) Proportion of Time Full-time Employed Mothers vs. Fathers in Two Parent Households Spent in Day (48 Hours Per Household)

(b) Proportion of Time Mother vs. Father in Two Parent Households are Contacted First By Type of External Decision Maker

Notes:
Panel (a) shows the proportion of time spent by male versus female respondents on different activities. Respondents are married adults, working full-time with children under 18 from the American Time Use Survey from the BLS years 2015-19 combined. There is a line at the equal time spent on an activity by male versus female respondents. The number at the top of each bar is the total hours spent on this activity by male and female respondents collectively (sums to close to 48 hours). For brevity we exclude some categories (e.g purchasing goods/services, caring for non-children, non-child related travel, and other activities). Full time working mothers tend to spend equal or more time on these excluded categories relative to the full time working fathers.
Panel (b) shows the proportion of time mothers and fathers are contacted by adult leaders who interact with parents. There is a line at the equal amounts of contact to mothers versus fathers. Respondents were 300 adults who interact with parents and self-identified as doing so mostly within a certain role (eg., Teacher, Nurse, Sports Leader) see Appendix K for details. We randomized whether respondents were asked the following question about a mother or a father: What proportion of the time do you contact the [father][mother] first if only contacting one parent first?

school-related requests. Schools therefore provide an ideal setting to investigate external demands for parental involvement by gender. To do so, we develop a theoretical model to inform the design of a field experiment in a K-12 school setting. Specifically, we send emails with phone numbers for both parents in a fictitious two-parent household to the near-universe of US school principals (N = 80,071), asking the principal to contact a parent by phone about a general school-related inquiry. We vary which parent sends the email and the information about their availability to disentangle whether discrimination stems from decision makers' beliefs about parents' responsiveness or from other deterrents. Beliefs about responsiveness might include the perception that women are more available because they are stay-at-home mothers or that women naturally want to be more involved in a school-related decision and will therefore be more responsive than men. Other deterrents might include distaste for calling a specific parent, systemic factors, social norms, or beliefs not related to responsiveness.

Experimentally varying the information about parents' availability allows us to investigate whether the gender gap can be mitigated by households changing the signals they send. As our experiment shows, signaling parental availability goes only so far in effecting change. Our model allows us to further explore other attributes at play, such as the prevailing gender norms of the decision makers, schools, and geographic locations. We show that such attributes impact the inequality in demands on parents' time, implying the gender gap might be mitigated through policies targeting behavioral change in specific sub-groups.[3]

We find striking gender and treatment differences. Principals are significantly more likely to call mothers first in the baseline treatment, which contains no signal about parents' availability. On average, mothers are called first 1.4 times more than fathers (59% versus 41%), providing direct and novel evidence of greater external demands on mothers in our setting. We believe that our findings are a first step toward documenting gender inequality in external demands, which are plentiful within the school setting (e.g., picking up a sick child, volunteering for school events) and beyond (e.g., which parent schedules doctor visits, who coordinates extracurriculars, and who grandparents expect to take care of a child's needs). Thus we suspect that our findings represent a lower bound on the overall external demands on mothers' versus fathers' time.

In addition to documenting this gender gap in external demand for parents' time, we explore the reasons it arises and test potential mechanisms. Specifically, we show that signaling that the father is more available mitigates the inequality and causes mothers to be called less than half the time. It is notable, however, that even when fathers signal that they are more available, mothers still get 26% of the calls. In contrast, signals that reinforce stereotypes that mothers are more available cause them to receive 90% of the calls. Strikingly, even when the email comes from the father and he signals his availability, 12% of the calls are still directed to mothers. This highlights an important asymmetry in the effectiveness of informational interventions in closing the observed gender gap in external demands for parents' time.

To identify the mechanisms underlying any differential demand for parental involvement, we pair a novel theoretical model with our field experiment and a survey. Our model shows

[3]The scope of this paper is exclusive to two-parent households with a male and female parent. We acknowledge that there are many types of households and more gender identities, but we believe that work using the two extreme ends of the gender spectrum (male/female) is an important first step in exploring how gender identity affects external demands on a person's time. However, we believe that exploring the effect of external demands in other settings is an important question for future work, and we discuss this in Appendix H.

how decision makers choose whether to contact a mother, father, or neither parent. It allows us to attribute any differences we find to statistical discrimination on the basis of beliefs about parents' responsiveness or to other factors, which we identify through a separate survey with school administrators. As mentioned earlier, these factors could include distaste, systemic factors, or beliefs unrelated to responsiveness. Our randomized signals about parents' availability impact a decision maker's beliefs only about the benefit of a call due to changes in availability. Thus, the differences in the proportions of calls across signals tell us what happens when those beliefs change. Any residual differences in the proportions of calls to mothers versus fathers are attributable to other deterrents.

Motivated by our survey evidence, we also use this flexible model to explore other belief-based channels—such as parental expertise, desire for involvement, or beliefs about mothers being stay at home parents—which we do in a separate set of treatments. We find that signaling that both parents want to be equally involved in the decision or that both parents work full time does not reduce the share of calls to mothers. This finding suggests that beliefs about parents' relative expertise, involvement, or employment status are not the primary drivers of the results.[4]

This paper extends the existing literature in four important ways. First, we experimentally document a novel gender gap in external demands for parental involvement. While prior research has found that women spend significantly more time on child-related tasks than men in two parent households, our study is the first to show that this inequality is in part driven by external demands for parental involvement. This gender inequality has considerable economic and social costs for women and men, who both report a desire for a more equal distribution of child-related tasks (Pew Research Center, 2015). In our own survey of parents in households with school-age children,[5] we find that women report being contacted by the school more often than men yet wish they were contacted less often, while men wish they were contacted more often. We also find that women are significantly more likely to be the point of contact for external decision makers across a wide range of child-related domains, from doctors' offices to extracurricular sports coaches to religious leaders (see panel (b) of Figure 1).[6] Perhaps most importantly, in our survey almost two-thirds of women self-report that these child related interruptions were something they considered when choosing

[4]The model also allows decision maker beliefs as well as other deterrents to vary by school characteristics. This variability allows us to study differences related to gender norms in Section 5.3.4.

[5]We detail the survey in Appendix K.

[6]Women, anticipating greater external demands for parental involvement long before having children, may be pushed toward more flexible jobs, leading to substantial labor market penalties, including reduced labor force participation (Kleven et al., 2021; Mas and Pallais, 2020; Bursztyn et al., 2017; Mas and Pallais, 2017; Pertold-Gebicka et al., 2016; Anderson et al., 2002) and curbed earnings (Cortes and Pan, 2021; Goldin, 2014; Gicheva, 2013).

their job, while less than one-third of men reported these same concerns.

Relatedly, prior research has documented the effects of childcare disruptions on women's labor market outcomes. Price and Wasserman (2022), for example, show that summer childcare constraints contribute to career choices and earnings for women with school-aged children, in line with findings from Duchini and Van Effenterre (2022). Similarly, the COVID-19 pandemic, and the associated school and daycare closures, led to significantly larger declines in women's employment and labor force participation relative to men. The negative effects have been especially large for mothers of school-aged children, leading to significant declines in their mental and physical health.[7] Understanding whether external demands for parental involvement contribute to gender inequalities in child-related tasks can shed light on the drivers of the persistent gender earnings gap and inform policies aimed at mitigating persistent gender inequalities.

Second, we contribute to the growing literature on the role of information in reducing discrimination. Prior work in economics and social psychology has considered the role of individual-specific information in reducing reliance on group statistics for evaluations (also known as statistical or belief-based discrimination). This literature has produced mixed evidence. While several recent studies show that providing accurate information reduces statistical discrimination (Laouénan and Rathelot, 2022; Bohren et al., 2019), others have found no discernible effects (Bertrand and Mullainathan, 2004; Oreopoulos, 2011).

Our paper advances this literature by documenting a striking asymmetry in the effect of information on reducing discrimination. In our field experiment, we test whether providing information about parents' availability mitigates the gender gap in external demands for parental involvement. Notably, while we find that signaling the availability of fathers moves calls away from mothers, we also document the limits of this informational intervention. Specifically, we find that signaling the high availability of mothers leads to mothers being contacted 90% of the time, while signaling the high availability of fathers increases calls to fathers only up to 74%.

A related literature to which we contribute investigates the underlying sources of discrimination. While field experiments lend themselves to identifying the existence of discrimination and its incidence, few experiments can identify the mechanisms that lead to discriminatory behavior (Bertrand and Duflo, 2017). The two most-studied mechanisms for discrimination in economics are tastes/preferences (Becker, 1957) and beliefs (Phelps, 1972;

[7] Cowan et al. (2023); Adams-Prassl et al. (2023); Couch et al. (2022); Garcia and Cowan (2022); Hansen et al. (2022); Amuedo-Dorantes et al. (2020); Zamarro and Prados (2021); Sevilla and Smith (2020); Montes et al. (2021); Heggeness (2020); Russell and Sun (2020); APA (2021).

Arrow, 1973; Aigner and Cain, 1977), with recent work emphasizing the importance of indirect discrimination stemming from systemic and institutional factors (Bohren et al., 2022; Kline et al., 2022). We join a small but growing literature that attempts to differentiate these sources of discriminatory behavior.

Prior research has employed field experiments to tease out the true sources of discriminatory behavior. For example, List (2004) examines racial discrimination in the baseball card market, Islam et al. (2018) investigate how patients choose a physician, and Bohren et al. (2019) examine gender discrimination in a mathematics forum. We advance this literature by using a simple, static theoretical model combined with a field experiment to identify separate parameters that capture the availability beliefs versus other deterrents which lead to discriminatory behavior.

Finally, this paper contributes to the literature on institutional, structural, or systemic discrimination. Prior work in sociology and economics has explored the idea that discrimination may be perpetuated by organizations or structures in addition to individuals (for discussions, see Small and Pager, 2020; Bohren et al., 2022; Kline et al., 2022; Scott, 2013; Council, 2004; Powell and DiMaggio, 2012). We provide novel evidence of systemic discrimination by showing that school principals' optimizing behavior ends up creating worse outcomes for some individuals in society and arguably for society as a whole. As Small and Pager (2020) argue, institutional discrimination deserves particular attention given the deeply ingrained nature of systemic practices and their long-lasting consequences.

Notably, the patterns that we document represent only a small share of the overall gender inequality in external demands for parental involvement. While the gender gap in school-related interruptions closely mirrors gender gaps in other child-related and household domains, this is only one of many settings where women are disproportionately more likely to experience child-related interruptions on a daily basis.[8] The gender inequality in physical housework, for example, has remained largely unchanged since the mid-1990s, with men spending about half as much time on housework as women in similar households (Bianchi et al., 2012). Furthermore, men's housework hours tend to be disproportionately allocated toward relatively infrequent and flexible tasks (e.g., home repairs or yard work), while women shoulder many of the recurring daily tasks (e.g., cooking and childcare) that cannot be put off to a convenient time (Bianchi et al., 2006). Moreover, research across social sciences has increasingly drawn attention to "invisible" forms of labor, including emotional

[8]In our own survey, we find that women are significantly more likely to be contacted by external decision makers across a wide range of child-related domains, from doctors' offices to extracurricular sports coaches to religion leaders (see panel (b) of Figure 1). Other studies have documented this pattern in larger samples (Wikle and Cullen, 2023; Bianchi et al., 2006; Boye, 2015; Daly and Groes, 2017; Daminger, 2019; **?**; Charmes, 2019).

and cognitive labor, being disproportionately shouldered by women.[9] While these inequalities are more difficult to measure directly, our findings shed light on potential policies aimed at mitigating these gender gaps.

2 Theoretical Framework

Our theoretical framework aims to show how a decision maker who interacts with a two-person, heterosexual couple decides which person to call upon for a task. In our specific field experiment, the decision maker is a school principal, and the task is a discussion about enrolling at the school. However, our theoretical model is flexible enough to be applied to different types of decision makers (e.g., doctors, dentists, school teachers, sports coaches, music teachers, summer camp directors, organized religion leaders) and different types of tasks (e.g., picking up a sick child, waiting in line to enroll in lessons, volunteering for career day or a bake sale, taking the team on an overnight trip). Furthermore, our model could apply outside of parenting tasks to test just about any type of demand on a two-person household (e.g., for elder care, interior-design projects, home renovations, retirement planning).

We lay out a simple economic structure in Section 2.1 to capture the decision-making behavior of school principals when contacting parents. In Section 2.2 we describe the random utility model we use to study this environment. We then explain in Section 2.3 how our experimental variation integrates with the random utility model. Section 2.4 shows how we use the model to identify and estimate its structural parameters, most notably the parameters for the principal's beliefs versus other deterrents. Section 2.5 outlines key testable hypotheses of interest.

The model is quite flexible and can be extended in several directions. In Section 2.6, we discuss robustness and add two extensions. One captures heterogeneity in the characteristics of principals, and the other shows that the model can also easily accommodate interdependence between principals' beliefs about male and female parents.

2.1 Economic Structure

School principals are the decision makers in our model; their alternatives are to call a male parent first (m), call a female parent first (f), or call neither parent (n). We index decision

[9] Daminger (2019); Offer (2014); Lee and Waite (2005).

makers by $i = 1, \ldots, N$. We take the experiment for a given decision maker to end when they choose an alternative $j = 1, \ldots, J$ or at our exogenously-determined experiment end date. The observables in our experiment are then (1) the choice $y_i \in J$ for each decision maker and (2) the characteristics of the alternatives x_i that are shown to each decision maker.

We assume that decision makers potentially face different costs, c_i, of making a phone call. For instance, some may have inferior technology or be busier than others. We also assume that they potentially perceive different benefits from choosing different alternatives. We further assume there are two components to these benefits: the decision maker's belief about the value of a response from an alternative parent and the deterrents they face to calling that alternative. We let r_{ij} denote decision maker i's subjective valuation of a response from alternative j, inclusive of the decision maker's assessment of the likelihood of a response,[10] and we let δ_{ij} denote other deterrents to calling alternative j. We assume that each decision maker i knows c_i and δ_{ij}, has beliefs over r_{ij} and is risk neutral.[11]

2.2 Random Utility Model

We construct a random utility model (McFadden, 1974) of decision maker behavior in which a decision maker's utility is the difference between the benefits and costs of calling alternative j. For the expected utility maximizer i, the expected utility of calling alternative j is defined as

$$U_{ij} = \mathbb{E}(r_{ij}) - \delta_{ij} - c_i, \tag{1}$$

where δ_{ij} is positive if factors other than availability beliefs on average deter decision maker i from calling alternative j. We think of δ_{ij} as a generalization of a distaste parameter, which includes distaste but also other factors not related to beliefs about availability such as social norms. This is our basic random utility formulation.

Because calling no one incurs no cost and provides no benefit, we take the utility of calling neither to be zero. This normalization will play a crucial role in identification because choice in this context is determined by differences in utility, not levels.

Under this normalization and in our context of the choice between calling either of two parents or calling neither, decision maker i calls neither parent if both $U_{i,m} < 0$ and $U_{i,f} < 0$; calls the female parent if $U_{i,f} \geq 0$ and $U_{i,m} \leq U_{i,f}$; and calls the male parent if $U_{i,m} \geq 0$ and

[10]We assume that the alternative who is called will have the relevant expertise. In Section 2.6.1, we add a separate, non-degenerate belief about expertise.

[11]In Appendix G.5, we discuss relaxing the assumption of risk neutrality.

$U_{i,f} < U_{i,m}$.[12] We can conceptually break their choice between the three alternatives into two parts. One is the decision of whether to make a call; the other is how to decide which parent to call if they are going to make a call.

The decision maker makes a call if and only if

$$\max\left\{\mathbb{E}(r_{i,f}) - \delta_{i,f},\, \mathbb{E}(r_{i,m}) - \delta_{i,m}\right\} \geq c_i.$$

If the decision maker makes a call, they call the female parent when

$$\mathbb{E}(r_{i,f} - r_{i,m}) \geq (\delta_{i,f} - \delta_{i,m}),$$

and they call the male parent when

$$\mathbb{E}(r_{i,m} - r_{i,f}) < (\delta_{i,m} - \delta_{i,f}).$$

Notice that the cost, c_i, does not affect the decision of which parent to call; the decision maker incurs the same cost regardless of which parent they call. The cost plays a central role in deciding whether to make a call, whereas the choice of which parent to call depends only on the differences in beliefs and other deterrents.

2.3 Experimental Manipulation of Beliefs

Consider an experimental manipulation that sends informative signals to decision maker i about the availability of either the female ($r_{i,f}$) or male parent ($r_{i,m}$). For simplicity, we assume all priors and signals are normally distributed. That is,

$$\bar{r}_j \sim \mathcal{N}(r_j, \omega_j^2), \quad x_{ij} \sim \mathcal{N}(r_j, \sigma^2), \quad j \in \{f, m\},$$

where $\bar{r}_j$ and ω_j^2 are the prior mean and variance common to all i. x_{ij} is a signal of the *true* responsiveness r_j of j that we send to i, and the signal variance is σ^2.

We assume that the priors for $r_{i,f}$ and $r_{i,m}$ are independent of each other and also of the distributions for the cost and other deterrents parameters. This implies that when we send a signal about one parent (female or male), it shifts only the belief about the parent for which the signal was sent and, further, does not impact the δ_{ij} or c_i.[13] Because of our assump-

[12] We break ties in favor of calling the female parent, but this has no impact in terms of the theory since utility is continuous.

[13] We relax this assumption in Section 2.6.2.

tion that decision makers are risk neutral, only the marginal means of this distribution are relevant for the expected utility and therefore decisions.

We then have decision maker i's posterior mean $\widetilde{r}_{ij}$ for the responsiveness of parent j as

$$\widetilde{r}_{ij} = \lambda_j \bar{r}_j + (1 - \lambda_j) x_{ij}, \quad \lambda_j = \frac{1/\omega_j^2}{1/\omega_j^2 + 1/\sigma^2}. \tag{2}$$

Letting w_{ij} be an indicator for sending i a signal regarding r_j, and recalling that x_{ij} is the signal's value, Equation 1 becomes

$$\begin{aligned} U_{ij} &= (1 - w_{ij})\bar{r}_j + w_{ij}\widetilde{r}_{ij}(x_{ij}) - \delta_{ij} - c_i \\ &= (1 - w_{ij})\bar{r}_j + w_{ij}\left[\lambda_j \bar{r}_j + (1 - \lambda_j) x_{ij}\right] - \delta_{ij} - c_i \\ &= \bar{r}_j - (1 - \lambda_j)\bar{r}_j w_{ij} + (1 - \lambda_j) w_{ij} x_{ij} - \delta_{ij} - c_i \end{aligned} \tag{3}$$

for $j \in \{f, m\}$. Recall that the utility of calling neither parent ($U_{i,n}$) is assumed to be zero.

Using $\bar{\delta}_j$ to denote the average value of δ_{ij} and c to denote the average value of c_i across the distribution of principals, Equation 3 can be written as

$$U_{ij} = \alpha_j + \eta_j w_{ij} + \gamma_j w_{ij} x_{ij} + \varepsilon_{ij}, \tag{4}$$

$$\alpha_j = \bar{r}_j - \bar{\delta}_j - c, \tag{5}$$

$$\eta_j = -(1 - \lambda_j)\bar{r}_j, \tag{6}$$

$$\gamma_j = 1 - \lambda_j, \tag{7}$$

$$\varepsilon_{ij} = (c - c_i) + (\bar{\delta}_j - \delta_{ij}). \tag{8}$$

The ε_{ij} are econometric errors and are mean zero because the average terms $\bar{\delta}_j$ and c are absorbed in the constant α_j. Importantly, the random assignment of x_{ij} and w_{ij} imply that they are independent of ε_{ij}.

We assume that the ε_{ij} are each distributed according to the standard Gumbel distribution, which implies that the error differences are distributed according to the standard logistic distribution. We next make the identification argument in terms of these econometric errors.

2.4 Identification of Reduced-Form and Structural Parameters

Identification is straightforward given the following elements of our setting and our model:

1. The random utility model provides structure for the relationship between benefits,

costs, and outcomes.

2. Calling neither outcome provides a clear normalization because it provides no benefits and incurs no costs.

3. Experimental randomization establishes that the regressors are not dependent on the outcome variable.

4. The assumption that errors are drawn from the logistic distribution leads to equations for the outcome probabilities that are closed form.

This would be a standard random utility model if our reduced-form parameters α_j, γ_j, and η_j did not vary across the j choices. Having intercepts and slopes that vary across alternatives is, however, crucial to learning about how the experimental manipulation impacts the choices of decision makers. Fortunately, the model's structure allows us to identify these intercepts and slopes.

Here we state the identification result and provide intuition for this result. All proofs are in Appendix G, with the proofs from this section in Appendix G.2.

Result 1. *Given the assumptions of Sections 2.1–2.3, the reduced-form parameters α_j, γ_j, and η_j are identified for $j \in \{f, m\}$.*

We identify the reduced-form parameters using ratios of the proportions of signal-outcome pairs. We denote the proportions as p_j^t. Subscripts indicate alternatives and superscripts indicate treatments $t \in \{b, l\boldsymbol{F}, h\boldsymbol{F}, l\boldsymbol{M}, h\boldsymbol{M}\}$, where b is the baseline treatment, treatment $l\boldsymbol{F}$ sends the low signal about the female parent, treatment $h\boldsymbol{F}$ sends the high signal about the female parent, treatment $l\boldsymbol{M}$ sends the low signal about the male parent, and treatment $h\boldsymbol{M}$ sends the high signal about the male parent. For example, $p_m^{l\boldsymbol{F}}$ is the proportion of decision makers who receive the low signal about female parent availability and then call the male parent.

Given the assumption that $\alpha_n = 0$, each α_j intercept is directly identified by comparing the proportion of decision makers who receive no signal and call parent j (p_j^b for $j \in \{f, m\}$) and the proportion who receive no signal and call neither parent (p_n^b). To separately identify γ_j and η_j, we need to create variation in the term $w_{ij}x_{ij}$, that is, the interaction of the indicator variable for whether a signal was sent (w_{ij}) and the value of the signal (x_{ij}). This variation must be distinct from the variation in w_{ij} alone. We achieve this by sending two values of the signal about each alternative j with known cardinal values. Specifically, we send both a positive signal and a negative signal about each parent and assume the values are 1 and -1.[14]

[14] For a discussion of the impact of the chosen scale of signals, see Section 2.6.4.

We now turn to the identification of the structural parameters $\bar{r}_j$, $\bar{\delta}_j$, c, and λ_j, where we have the following result.

Result 2. *Given the assumptions of Sections 2.1–2.3 and Result 1, the structural parameters $\lambda_f, \lambda_m, \bar{r}_f, \bar{r}_m$, and $\bar{\delta}_m - \bar{\delta}_f$ are identified.*

Since we can identify γ_j, Equation 7 provides for the identification of λ_j, which is the weight that decision makers place on their prior belief when updating. Given η_j and λ_j, Equation 6 allows us to identify the prior belief $\bar{r}_j$. Combining $\bar{r}_j$ and α_j, Equation 5 identifies $\bar{\delta}_j + c$. We can then combine $\bar{\delta}_f + c$ and $\bar{\delta}_m + c$ to identify the difference in the other deterrents parameters for calling female versus male parents, $\bar{\delta}_f - \bar{\delta}_m$.

We can develop intuition by looking at the relationships between the reduced-form and structural parameters as given in the following three equations:

$$\begin{aligned} \bar{r}_j &= -\frac{\eta_j}{\gamma_j}, \\ \bar{\delta}_m - \bar{\delta}_f &= \frac{\eta_f}{\gamma_f} - \frac{\eta_m}{\gamma_m} + \alpha_f - \alpha_m, \\ \lambda_j &= 1 - \gamma_j. \end{aligned}$$

The second equation is derived from the definition of α_j and can be expressed as

$$\alpha_f - \alpha_m = \bar{\delta}_m - \bar{\delta}_f + \bar{r}_f - \bar{r}_m. \tag{9}$$

We can interpret this as indicating that the magnitude of the gender inequality (if indeed $\alpha_f > \alpha_m$) derives from the excess deterrents decision makers face for calling male parents plus their excess belief in the availability of female parents.

2.5 Testable Hypotheses

With both the reduced-form and structural parameters identified, we can now state the following testable implications of the theory.

Hypothesis 1. *There is a gender inequality in external demands for parental involvement. To detect this inequality, we need to see $p^b_f > p^b_m$; that is, the proportion of decision makers who receive no signal and call the female parent is larger than the proportion who receive no signal and call the male parent. This is equivalent to $\alpha_f > \alpha_m$ in terms of the reduced-form parameters and $\bar{r}_f - \bar{\delta}_f > \bar{r}_m - \bar{\delta}_m$ in terms of the structural parameters.*

Hypothesis 2. *Without intervention, decision makers believe that female parents are more available than male parents. We find support for this hypothesis if* $\bar{r}_f > \bar{r}_m$*.*

Hypothesis 3. *On average, the deterrents decision makers face to calling male parents is larger than the deterrents to calling female parents. We find support for this hypothesis if* $\bar{\delta}_m - \bar{\delta}_f > 0$*.*

We can combine the testable implications in Hypotheses 1–3 with Equation 9 to say the following about the sources of differential treatment of female and male parents.

Result 3 (Sources of differential treatment)**.** *If there is gender inequality [Hypothesis 1] and*

1. *Hypothesis 2 is supported (i.e.,* $\bar{r}_f > \bar{r}_m$*) while Hypothesis 3 is not (i.e.,* $\bar{\delta}_m - \bar{\delta}_f \leq 0$*), then differential treatment stems from beliefs about availability.*
2. *Hypothesis 3 is supported (i.e.,* $\bar{\delta}_m - \bar{\delta}_f > 0$*) while Hypothesis 2 is not (i.e.,* $\bar{r}_f \leq \bar{r}_m$*), then differential treatment stems from other deterrents to calling male parents.*
3. *both Hypothesis 2 (i.e.,* $\bar{r}_f > \bar{r}_m$*) and Hypothesis 3 (i.e.,* $\bar{\delta}_m - \bar{\delta}_f > 0$*) are supported, then differential treatment stems from both beliefs and other deterrents.*

The parameters required to test Hypotheses 1–3 are identified from data as demonstrated in Result 2.

2.6 Model Extensions and Robustness

2.6.1 Beliefs about Both Availability and Expertise

Until now, we have assumed that decision maker beliefs about the value of calling parents only incorporate the parents' availability. We can expand our conceptualization of decision maker beliefs to also include the expertise of the parents. If we model these two components of beliefs as multiplicative, that is, $\mathbb{E}[r_j q_j] = \bar{r}_j \bar{q}_j$, decision maker utility in Equation 1 becomes

$$\mathbb{E}(U_{ij}) = \bar{r}_j \bar{q}_j - \delta_j - c_i.$$

We now reexamine the identification of our structural parameters given this new element of the model. To do so, we must take a stand on how decision makers will interpret our signals about availability now that their beliefs also contain the expertise component.

If the signal about availability does not impact the belief about expertise, the prior belief about expertise is simply carried along with the signal so that the updated belief is

$$\widetilde{qr}_{ij} = \lambda_j \bar{q}_j \bar{r}_j + (1 - \lambda_j) \bar{q}_j x_{ij}$$

and the expected utility after updating on the signal is

$$\begin{aligned}\mathbb{E}(U_{ij}) &= (1-w_{ij})\bar{q}_j\bar{r}_j + w_{ij}\widetilde{qr}_j(x_{ij}) - (\delta_j + c_i)\\ &= \bar{q}_j\bar{r}_j - (1-\lambda_j)\bar{q}_j\bar{r}_j w_{ij} + (1-\lambda_j)\bar{q}_j w_{ij}x_{ij} - (\delta_j + c_i).\end{aligned}$$

The following equations map the reduced-form parameters to the structural parameters:

$$\begin{aligned}\alpha_j &= \bar{q}_j\bar{r}_j - \bar{\delta}_j - c,\\ \eta_j &= -(1-\lambda_j)\bar{q}_j\bar{r}_j,\\ \gamma_j &= \left(1-\lambda_j\right)\bar{q}_j,\end{aligned}$$

and we have $\eta_j = -\gamma_j\bar{r}_j \Leftrightarrow \bar{r}_j = -\frac{\eta_j}{\gamma_j}$ as in the base model. That is, our experimental variation continues to identify the availability belief even when the belief contains more than just availability. However, we no longer cleanly identify the updating or "other deterrents" parameters. Instead, we have $\lambda_j = \frac{1-\gamma_j}{\bar{q}_j}$ and $\bar{\delta}_j + c = -\bar{q}_j\frac{\eta_j}{\gamma_j} - \alpha_j$. Both are polluted by the average belief about expertise. Moreover, the distortion depends on the both the sign and the magnitude of the average belief about expertise.

To address this concern, we introduce a variation on our main set of five treatments (henceforth "main" variation), which we term the "equal decision" variation. In the equal decision variation, we send the same set of five messages, but to each we add the statement "This is the type of decision we both want to be involved in equally" to fix the decision maker's belief about parental expertise.

If we do not take a stand on the value of the expertise signal and label that value $q_j^{'}$, the updated belief becomes

$$q_j^{'}\widetilde{r}_{ij} = \lambda_j q_j^{'}\bar{r}_j + (1-\lambda_j)q_j^{'}x_{ij}.$$

All five treatments, including the baseline, receive this same message $q_j^{'}$ about expertise, so it appears in both terms on the right-hand side. Similar to the main variation discussed above, we once again can identify the beliefs about availability as $\bar{r}_j = -\frac{\eta_j}{\gamma_j}$, but we do not get $\bar{\delta}_j + c$ cleanly; instead we have $\bar{\delta}_j + c = -q_j^{'}\frac{\eta_j}{\gamma_j} - \alpha_j$.

If we are willing to assume that $q_j^{'} = 1$ (the same value we have assumed for positive signals about availability), then we cleanly identify the other deterrents + cost term as the same combination of reduced-form parameters as in the main variation where we assume there is no belief about expertise. In fact, if we are willing to assume any particular value for the expertise signal, we can cleanly identify the other deterrents + cost term in this more general case where beliefs include an expertise component.

We can actually do more to understand beliefs about expertise. If we combine the baseline treatment from the main variation with the remaining four treatments from the equal decision variation with $q_j^{'} = 1$, we have

$$\widetilde{qr}_{ij} = \lambda_j \bar{q}_j \bar{r}_j + (1 - \lambda_j) x_{ij} q_j^{'} = \lambda_j \bar{q}_j \bar{r}_j + (1 - \lambda_j) x_{ij}.$$

The mapping from reduced-form to structural parameters becomes

$$\begin{aligned}
\alpha_j &= \bar{q}_j \bar{r}_j - \bar{\delta}_j - c, \\
\eta_j &= -(1 - \lambda_j) \bar{q}_j \bar{r}_j, \\
\gamma_j &= \left(1 - \lambda_j\right).
\end{aligned}$$

Similar to the main variation that ignores beliefs about expertise, γ_j identifies λ_j, and we can recover both $\bar{q}_j \bar{r}_j = -\frac{\eta_j}{\gamma_j}$ and $\bar{\delta}_j + c = -\frac{\eta_j}{\gamma_j} - \alpha_j$. The only difference is that the belief now encompasses expertise. Given enough statistical power, we can then divide $\bar{q}_j \bar{r}_j$ by the $\bar{r}_j$ identified in the main variation to recover $\bar{q}_j$ separately.

2.6.2 Adding Decision Maker Characteristics

Until now, we have assumed that all decision makers are identical in terms of their observable characteristics. We can, however, easily allow for decision makers to differ in their beliefs and tastes according to an observable characteristic; we are especially interested in whether the decision-maker works at a religious school as this may correlate with holding more traditional gender normative views. To be clear, we do not change the signals that we send to principals in any way. This model extension simply allows the signals we send to impact the beliefs of different types of decision makers differently. Appendix G.3 contains the details of this model extension, the identification result, and additional testable hypotheses.

2.6.3 Relaxing the Independence Assumption

Now suppose that a signal about one parent induces the decision maker to update their belief about both parents.[15] This could happen, for instance, if the decision maker's beliefs about the parents are correlated or if the decision maker directly infers information about

[15] We present the theory for the case where decision makers are not differentiated by characteristics as in Section 2.4. It is straightforward to combine the two extensions to have both the decision maker characteristics and the cross impact of signals.

both parents from a signal about just one parent.

This model extension allows us to identify the magnitude of the impact of signals about one parent on the belief about the other parent. It also allows us to determine whether decision makers put different weight on their prior beliefs when signals are about women versus men. The generalized utility formulation and mapping to reduced-form and structural parameters is available in Appendix G.4, along with the identification result and an additional testable hypothesis.

Importantly, all empirical results below allow for, and can quantify, the interdependence of beliefs about the two parents that is discussed in Appendix G.4.

2.6.4 Signal Values and Scaling

We have so far assumed that decision makers take the value of any positive signal to be $x_{ij} = 1$ and the value of any negative signal to be $x_{ij} = -1$. If we change the assumed values of the signal symmetrically (e.g., both change from magnitude 1 to magnitude 2), η_j does not change but γ_j does. The intuition is as follows: we have not changed whether a signal arrives or not, so the impact of receiving any signal (i.e., η_j) does not change. However, although the signal's value is now assumed to be different, the term $(1 - \lambda_j)w_{ij}x_{ij}$ in Equation 3 does not vary with our assumption about the value of x_{ij}. Instead, when we change x_{ij}, the value of $\gamma_j = (1 - \lambda_j)$ adjusts to compensate since w_{ij} is simply an indicator for whether any signal is sent. Therefore γ_j is scaled in the opposite direction of the signal value. For instance, if the signals go from magnitude 1 to magnitude 2, γ_j is cut in half. The intercepts, α_j, do not change since they are entirely determined by the baseline.

If we change the assumed value of just one of the signals (e.g., to $+2/-1$ or $+1/-2$), the new γ_j falls between the γ_j for the $+1/-1$ and $+2/-2$ cases. η_j also changes, falling when the positive signal is larger and rising when the negative signal is larger. Any of these changes then ripple through to the structural parameters.[16] In short, as long as we are willing to take a stand on the value of the signals, the structural parameters are identified. However, the identified values of the structural parameters depend on the values we posit for the signals.

[16]Note that the treatment effects parameters would *not* change as long as they are still just dummy variables for each treatment. The simple relationships between the treatment effects and the reduced parameters would be modified to account for the ignored value of the signals in the treatment effects model.

3 Field Experiment

Our theoretical model and a survey of school administrators inform the design of a large-scale field experiment, which consists of sending email messages to a near-universe of US school principals. The emails are sent from a set of fictitious parents, one male and one female.[17] Email is a common way for parents to contact schools; our own survey found that three-fourths of educators report being contacted by parents via email at least once a month.[18] Additionally, several recent studies have used emailing schools as part of their methodology to document discrimination against students with disabilities, of certain races, or with homosexual parents (see, for example, Diaz-Serrano and Meix-Llop (2016); Bergman and McFarlin Jr (2018); Ahmed et al. (2020); Oberfield and Incantalupo (2021); Cantet et al. (2022); Hermes et al. (2023)). In the study most like our own, Hermes et al. (2023) email childcare centers in Germany from either the mother or the father and find that response rates are similar but responses to mothers are shorter and less positive than responses to fathers.

3.1 Setting

Our experiment takes place in a K-12 school setting. A large portion of the general population, about 40% of households in the US, have school-aged children (NCES, 2021), and 97% of parents send their children to school outside the home (NCES, 2021). Schools are an ideal setting to explore external demands on parents' time because the gender gap in time spent on children in school-related activities closely mirrors the overall tendency for mothers to engage in more child-related tasks than fathers (BLS, 2021).

For several reasons, we believe that any gender gaps that we document in our specific task will generalize to other tasks in the school setting, such as picking up a sick child, volunteering for the book fair, or joining the Parent Teacher Association (PTA). First, educators in our survey say that they would contact the mother first in many of these scenarios (we discuss the survey in Appendix K). Second, the gender distribution of these tasks is significantly skewed; mothers comprise almost 90% of PTA members, and only 13% of fathers report high levels of involvement in their child's school activities, compared to 53% of mothers.[19]

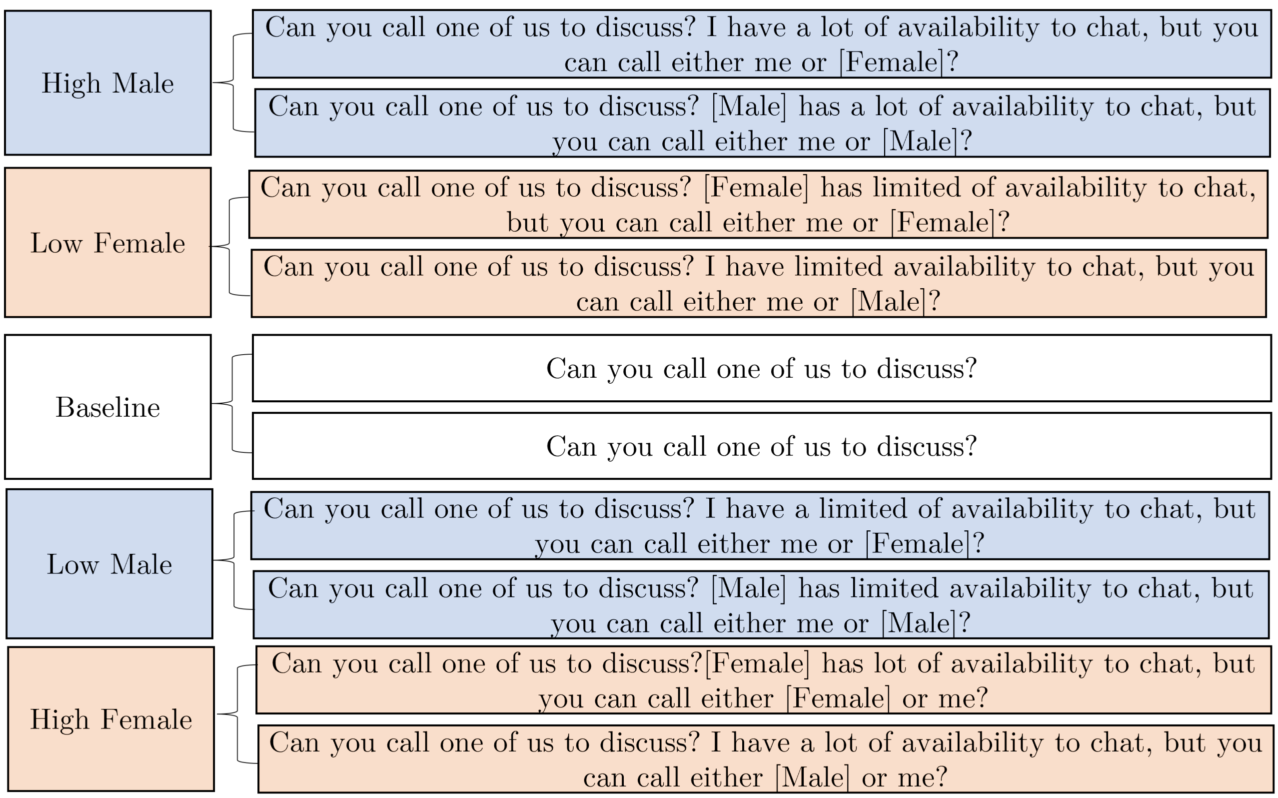
[17]We acknowledge that there are many different types of households and more than two genders, and we discuss this further in Appendix H. We describe our data collection process in more detail in Appendix J as well as some of the ethical considerations in Appendix I.

[18]We discuss the survey in detail in Appendix K.

[19]See Daly and Groes (2017); Belkin (2009); Scottland (2020).

Furthermore, although the gender gap in external demands for parental involvement is established for our test case of outreach from a school administrator, we expect that it is indicative of a dynamic that is likely present in a wide range of social situations that require parental attention or input. As shown in Figure 1, mothers spend more time on many tasks than fathers, and decision makers from a variety of places beyond school self-report contacting mothers more than fathers. If these other inequalities are also partially driven by external demands, our findings likely represent a lower bound for the overall gender gaps in external demands for parental involvement.

Figure 2: Field Experiment Variation in Messages



Notes: In this figure we show pertinent portion of variation in the messages we sent to schools. The parent who sent the email always had their phone number listed first. Above we show the message sent from the male parent (cc'ing the female parent) then the message from the female parent (cc'ing the male parent). The full text of example email messages is available in Appendix Section F

3.2 Messages

In our experiment, school principals receive an email from a fictitious two-parent, heterosexual household. The email states that the parents are searching for a school for their child and would like to have a phone discussion about it. We provide separate phone numbers for each parent. The email sender's phone number is always listed first, and we randomize whether the primary sender is the father or mother. We developed the specific message in consultation with school administrators from a variety of schools (public, private, and charter). Our conversations and survey evidence (Appendix K) confirmed parents frequently make general email inquiries to schools before enrolling and that it is common for one parent to email, copying the other parent.

We then augment our baseline message across our other treatments by adding a sentence indicating the availability of a specific parent in the two-parent household. Figure 2 shows examples of the exact variation in wording. Details of the exact names and email addresses used in the experiment are in Appendix J, and the full text of the messages is in Appendix F.

We designed these messages based on our theoretical model discussed in Section 2 as well as a survey we conducted with school administrators detailed in Appendix K. Our survey findings revealed that a key dimension on which school administrators could be statistically discriminating is differential beliefs about mothers' relative availability. Specifically, a common reason administrators gave for calling mothers first was "I expect this person to be more likely to respond quickly." One of the model's key results is that by varying the strength (low/high) of the signals about each of our parents' availability, we can disentangle the extent to which the gender inequality is driven by beliefs about mothers having higher availability versus other deterrents.

3.3 Sample Frames and Data Collection

During the summer of 2022[20], we sent emails to a near-universe (a sample of 80,071) of school principals across the US. We begin by describing the main variation of our experiment, which was sent to 30,471 school principals. We observe whether any call is made to any of the phone numbers we list including phone calls where no voicemail was left. We also know the precise time, date, content, and length of any voicemail left for our parents. We use this information to match each phone call back to the original decision maker who received one of our treatment emails. Appendix J provides more details about the experimental design, data collection, and matching process.

Approximately two weeks after we sent the initial email, we sent a second email telling the decision maker we no longer needed to speak with them, thus releasing them from any obligation to continue trying to reach us. The vast majority of calls from principals are made

[20]Over the course of 2021, we conducted a series of pilots with a total of 3,267 observations to iron out implementation logistics. Some pilots were sent out during the school year, while others during the summer. Notably, we did not observe significant differences in response rates by time of year.

within the first week of the original email being sent.

4 Empirical Strategy

Our main outcome of interest is whether a decision maker calls the female parent, the male parent, or neither parent. To test whether our treatments have any effect on the relative proportions of no call, calling the female parent first, or calling the male parent first, we run the following multinomial logit regression:

$$p_{ij}(x) = \frac{e^{\beta_j^{l\boldsymbol{M}}(LowMale)+\beta_j^{h\boldsymbol{M}}(HighMale)+\beta_j^{l\boldsymbol{F}}(LowFemale)+\beta_j^{h\boldsymbol{F}}(HighFemale)+\alpha X_i}}{\sum_{k\in n,f,m} e^{\beta_k^{l\boldsymbol{M}}(LowMale)+\beta_k^{h\boldsymbol{M}}(HighMale)+\beta_k^{l\boldsymbol{F}}(LowFemale)+\beta_k^{h\boldsymbol{F}}(HighFemale)+\alpha X_i}}. \tag{10}$$

In this regression model, p_{ij} is the probability that individual i calls neither parent ($j = N$), the female parent ($j = F$) or the male parent ($j = M$). We next have treatment indicators for each of the non-baseline treatments: LowMale, HighMale, LowFemale, and HighFemale. We can also include a vector X_i of covariates including which parent the email was sent from (cc'ing the other parent) and attributes of the decision maker and their school.

In subsequent analysis, we let the outcome variable instead be binary taking the value one when a female parent is called and zero otherwise. We then run a simple linear regression for ease of interpreting the coefficients.

4.1 Mapping Treatment Effects to Reduced-Form and Structural Parameters

When we do not include the vector of covariates X_i, it is straightforward to map the coefficients from the treatment effects regression in Equation 10 to the reduced-form parameters from Equation 4.[21] This equation is developed in Appendix G.2 and is reproduced in the next paragraph.

We run an unordered logit over decision maker i's choice to call neither parent (n), the female parent (f), or the male parent (m). Taking calling neither parent as the baseline, we have the following equation for calling the female parent:

[21] Further, we can map the treatment effects to the more general reduced-form equation that includes impacts of signals on the beliefs about both parents in Equation 28 in Appendix 2.

$$p_{if}(x) = \frac{e^{\alpha_f + \eta_f^F w_{i,f} + \eta_f^M w_{i,m} + \gamma_f^F w_{i,f} x_{i,f} + \gamma_f^M w_{i,m} x_{i,m}}}{1 + \sum_{k \in f,m} e^{\alpha_k + \eta_k^F w_{i,f} + \eta_k^M w_{i,m} + \gamma_k^F w_{i,f} x_{i,f} + \gamma_k^M w_{i,m} x_{i,m}}}.$$

We also have the analogous equation for calling the male parent.

Notice that it matters both which parent is called and which parent the message is about. η_f^F captures the impact of a signal about the female parent on the probability of calling the female parent, while η_f^M captures the impact of a signal about the male parent on the probability of calling the female parent.

The mapping from the reduced-form coefficients to the treatment effects coefficients is simple and intuitive. To be concrete, let's look at the impact of signals about the male parent on the probability of calling the female parent. The reduced-form equation separates this effect into the impact of sending any signal and the impact of the signal's value, which we assume to be 1 or -1. The treatment effects equation separates this effect into the impact of the high signal about the male parent and the impact of the low signal about the male parent. Thus we have $\beta_f^{h\boldsymbol{M}} = \eta_f^M + \gamma_f^M$; that is, the treatment effect from the high signal about the male parent is equivalent to adding together the impact of receiving any signal about the male parent and the impact of the signal value being 1. Similarly, $\beta_f^{l\boldsymbol{M}} = \eta_f^M - \gamma_f^M$; that is, the treatment effect from the low signal about the male parent is equivalent to adding together the impact of receiving any signal and the impact of the signal value being –1.

The same relationship holds for each combination of parent called and signal sent, that is, signals about the female parent and the probability of calling the female parent, signals about the female parent and the probability of calling the male parent, and signals about the male parent and the probability of calling the male parent. The two regressions simply decompose the effects of the signals about the male parent in different ways.

To build intuition, we show here how to map the treatment effects and reduced-form parameters to the proportions of decision makers in the relevant outcome-signal pairs. For instance,

$$\beta_f^{h\boldsymbol{M}} = \eta_f^M + \gamma_f^M = \ln p_f^{h\boldsymbol{M}} - \ln p_n^{h\boldsymbol{M}} - (\ln p_f^b - \ln p_n^b).$$

That is, the impact of the high signal about the male parent on calls to female parents is determined by calculating the difference in log proportions of calls made to female parents versus no parents under the high signal about male parents, as compared to the baseline log difference of calls to female parents versus male parents.[22]

[22] The analogous relationship for the low signal is $\beta_f^{l\boldsymbol{M}} = \eta_f^M - \gamma_f^M = \ln p_f^{l\boldsymbol{M}} - \ln p_n^{l\boldsymbol{M}} - (\ln p_f^b - \ln p_n^b)$.

5 Results

We are balanced on observable variables across our treatments as shown in Table C.1. Although we had intended to send an equal number of emails from fathers and mothers as well as an equal number of emails in each of our treatments, these design choices were not attained due to some computing errors.[23] Our main results are based on reweighted data such that there is balance in the number of messages sent in each of our five messages (Figure 2), and there is balance between the number of messages sent from fathers versus mothers within a treatment arm. However, our results are quantitatively and qualitatively the same when we randomly exclude observations to achieve balance as shown in Appendix L.

5.1 Gender Inequality with No Signal

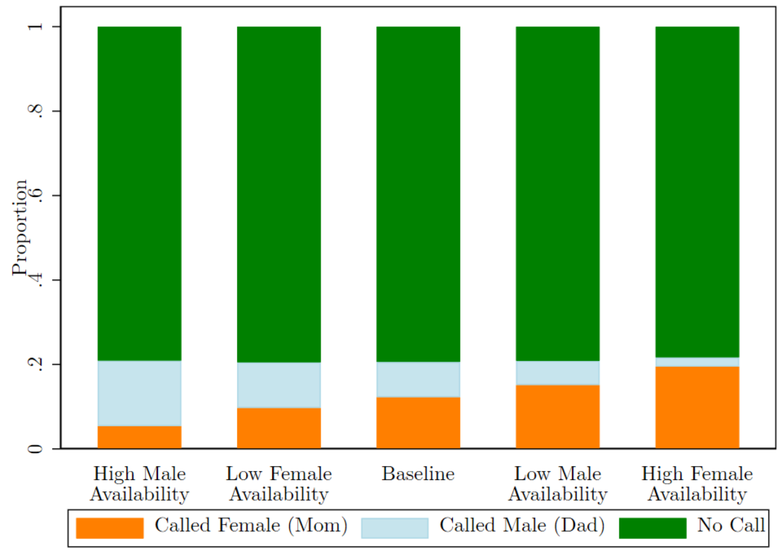
Table 1 and Figure 3 include information about the proportion of actions taken by decision makers in all of our conditions including our baseline condition, when there is no information about parents' availability. If there was no gender inequality and decision makers were randomly choosing which parent to call, we would expect the same proportion of calls to male and female parents. In line with Hypothesis 1 outlined in Section 2, we observe that 12% of school principals call mothers first, while only 8% call fathers first. The remaining 79% of decision makers do not call either parent.[24] The difference in calls to male and female parents is large and statistically significant ($Pr(T > t) = 0.00$). Thus we observe a clear gender gap when no signals are given to decision makers, with mothers being significantly more likely than fathers to be called first.

Another way to see this bias toward calling female parents is in the ratio of female-to-male calls, which is 1.4. This is well above the ratio of 1 that we would expect if decision makers were randomizing which parent to call, and it means that mothers are 1.4 times more likely than fathers to receive a call. Conditional on receiving a call back, mothers are called first 59% of the time in baseline.

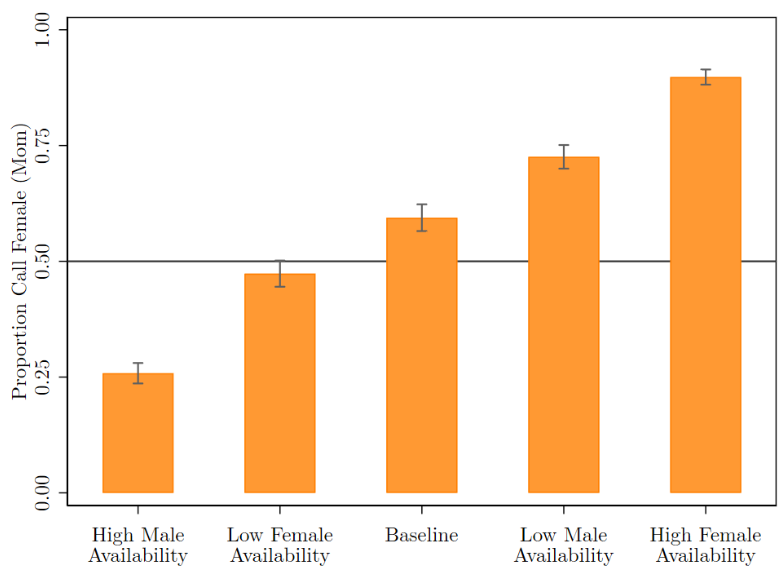
[23]The issue arose due to the use of the "set seed" command in Stata but was not detected until after our experiment had been fully run. We have no reason to believe that this computing error has introduced any systematic bias into our results.

[24]A response rate of 21% seems in line with previous work. Recent studies where job applicants submit applications with a phone number and email to an employer find response rates from employers of about 8% to 11% (Agan and Starr, 2018). For work on a similar subject pool of school principals, in line with our expectations, the response rate by phone is lower than the response rate via email observed by others, which ranges from 40% to 70% (Diaz-Serrano and Meix-Llop, 2016; Bergman and McFarlin Jr, 2018; Ahmed et al., 2020; Oberfield and Incantalupo, 2021; Cantet et al., 2022; Hermes et al., 2023). Another related outcome is whether principals take a survey in response to an email request, where recent work finds only 14% of principals take this action (Neal et al., 2020).

Figure 3: Outcomes by Treatment



(a) All Outcomes



(b) Outcomes Conditional On Calling

Notes: In this figure we show the proportion of decision makers choosing to make no call, call the female parent (mom) or the male parent (dad) by the message sent to the decision maker in our Main variation. Panel (a) represents three outcomes from 30,471 decision makers, while panel (b) shows only the choices of those who made a phone call to at least one parent ($N = 7,778$). In Panel B we regress dummy variables for our five messages on a binary variable for whether the female parent was called first or the male parent. One-way t-tests comparing No Call, Call Female and Call Male are all statistically significant at the 5% level or below. Observations are weighted so that 50% of emails come from a female parent and 50% from a male parent (always CCing the other parent). See Table 1 for sample size by message and standard errors. See Figure B.2 for total number of no calls, calls to female parent or calls to male parent by message.

Table 1: **Summary Statistics by Treatment in Main Variation**

			Panel A: All Outcomes		
	(1)	(2)	(3)	(4)	(5)
	High Male (Hm)	Low Female (Lf)	Baseline (b)	Low Male (Lm)	High Female (Hf)
Called Female	0.05	0.10	0.12	0.15	0.19
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Called Male	0.16	0.11	0.08	0.06	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
No Call	0.79	0.79	0.79	0.79	0.78
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	7075	5931	5612	5700	6153
			Panel B: Conditional on Calling		
Called Female | Call	0.26	0.47	0.59	0.73	0.90
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Called Male | Call	0.74	0.53	0.41	0.27	0.10
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	1483	1216	1158	1190	1335

Notes: Standard errors are in parentheses. Observations are weighted so that there is 50% of emails from a female parent and 50% from a male parent, and so that all message types have equal weighting.

We suspect that the gender gap we document is a lower bound on the overall gender inequality in external demands for several reasons. First, the type of inquiry in our messages is not a stereotypical male or female question. We would expect external decision makers to exhibit an even stronger bias toward calling female parents if they needed to call a parent to pick up a sick child, discuss allergies, or help with a bake sale. Second, the school setting itself is universal and should not be perceived as a stereotypical male or female domain. We would expect even more inequality toward mothers if we had sent this same type of message to a doctor's office or to a dance school because health and dance are stereotypical female domains. In contrast, if we were to send an inquiry about joining a hockey league or about additional school fees, we might expect less of a bias toward mothers because both are stereotypical male domains.

We explore differences by domain in Section 5.4. However, joining an extracurricular team or paying additional fees (especially at a public school) are not as universal as the experience of being called to pick up a sick child. Furthermore, picking up a sick child is usually an unexpected event that causes a large interruption to a person's day, in contrast to less time-intensive and more flexible requests about an extracurricular team or school fees. As such, we believe that the inequality that we document in our setting is a lower bound on the inequality in external demands on mothers versus fathers when there is no signal about which parent to contact.

While our main analysis focuses on the first call, we find similar patterns when investigating multiple calls made by the same principals (Figure B.2). Conditional on calling, 52% of the principals in our sample make more than one call, with an average principal making 1.7

calls. Principals who make only one call are far more likely to call the mother than the father (64% to mothers versus 36% to fathers). For those who make two or more calls, only 41% of those who call the mother first then try the father, while 53% of those who call the father first then try the mother. The rate of two calls to the mother in a row is double the rate of two calls to the father in a row. Overall, this strongly supports our baseline finding that women are disproportionately more likely to field child-related external demands.

5.2 Impact of Signals on Gender Inequality

5.2.1 Explicit Signals about Availability

Figure 3 shows the proportion of calls made to female and male parents alongside no calls in panel (a) and conditional on a call being made in panel (b). It is clear from the figure that the signals about high and low availability change which parent receives a phone call and can either increase or decrease the baseline bias toward calling female parents.

To rigorously assess the effects of our messages with signals on bias toward calling mothers in comparison to the baseline message, Figure B.1 visually represents the outcomes from a multinomial logit model like that in Equation 10 (see Table A.1 for more details and this same model with and without control variables included). This same multinomial logit model allows us to decompose the mechanisms for the gender inequality into discrimination based on beliefs about availability versus other deterrents, which we discuss in Section 5.3.

Recall that we randomly vary signals about availability across four messages: HighMale, LowMale, HighFemale, and LowFemale. Two of these messages (HighMale and LowFemale) go against preexisting gender norms by stating that the father has a lot of availability or the mother has limited availability. Figures 3 and B.1 show that these messages cause calls to move away from mothers and toward fathers, which mitigates the gender gap in external demands. The HighMale message reverses the inequality so that mothers are now called 26% of the time, while the LowFemale message moves mothers and fathers close to parity, with mothers getting 47% of the calls and fathers the remaining 53% (Table 1).

In contrast, the remaining two messages, LowMale and HighFemale, affirm the gender norm that mothers are more available than fathers. We find that they exacerbate the existing inequality by pushing calls toward mothers and away from fathers. Specifically, stating that the father has low availability results in mothers being called 73% of the time, representing a 24% increase in calls to mothers from the baseline. This change is almost symmetric to the 20% decline in calls to mothers from baseline caused by the LowFemale treatment.

Our results also highlight a striking asymmetry in the effect of our informational interventions. Notably, the HighFemale message stating the mother has high availability results in her being called almost 90% of the time, which is in contrast to fathers getting at most about 74% of the calls under the HighMale message. Thus, there appears to be a ceiling on how much the father can become the primary point person for external demands, while no such ceiling exists for demands on mothers.

In general, our messages about low availability have smaller effects than those about high availability. It is possible that our messages, especially the signals about low availability, might be impacting principals' response rates. We check whether there is any variation in the no-call rate across our treatments and find that all of them result in a similar no-call rate between 78% and 79% (Table 1 and Figure B.1).

5.2.2 Nonverbal Signals

In our experiment we randomly vary verbal cues about which parent is more or less available. Our messages have large effects, with the HighFemale message resulting in 19% of principals calling the mother versus the HighMale message having only 5% calling the mother. This is a 14 percentage point difference, which reverses the gender inequality in favor of men (Table 1). However, there are also nonverbal cues that households can use to signal which parent is the primary point of contact. In our study we randomly assign whether an email comes from the female parent with the male parent cc'd or vice versa. The person sending the email is a nonverbal signal of which parent to contact first.

Pooling across our treatment messages in the main variation, we find that the no-call rate is similar for both types of senders, suggesting that principals are as likely to respond to an email regardless of the sender's identity (see Table 2). However, whether the email is sent by the mother or the father significantly impacts the gender gap in response. Specifically, sending an email from the mother results in the principal calling her 18% of the time and calling the father only 4% of the time, a 14 percentage point difference (similar to what we see between our HighFemale messages, where the mother is called 19% of the time, and HighMale messages, where the mother is called 5% of the time). In contrast, sending the email from the father results in the principal calling him 13% of the time and calling the mother only 7% of the time, a 6 percentage point difference (smaller than the difference between our HighFemale and HighMale messages). It is clear that the sender's identity has a large positive effect on who gets the first call. However, that effect is not symmetric for mothers and fathers.

Table 2: **Summary Statistics by Primary Email Sender**

	(1) All Msgs.	(2) High Male (Hm)	(3) Low Female (Lf)	(4) Baseline (b)	(5) Low Male (Lm)	(6) High Female (Hf)
		Panel A.i: Email Sent by Mother cc'ing Father For All Outcomes				
Called Female	0.18	0.08	0.18	0.20	0.21	0.21
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Called Male	0.04	0.13	0.03	0.01	0.01	0.01
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
No Call	0.79	0.78	0.79	0.79	0.78	0.79
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	15560	3712	2726	3108	2895	3119
		Panel A.ii: Email Sent by Mother cc'ing Father Conditional On Calling				
Called Female | Call	0.83	0.39	0.86	0.98	0.96	0.97
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Called Male | Call	0.17	0.61	0.14	0.02	0.04	0.03
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	3300	801	567	647	626	659
		Panel B.i: Email Sent by Father cc'ing Mother For All Outcomes				
Called Female	0.07	0.02	0.02	0.04	0.10	0.18
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
Called Male	0.13	0.18	0.19	0.16	0.11	0.04
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
No Call	0.79	0.80	0.80	0.80	0.80	0.78
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	14911	3363	3205	2504	2805	3034
		Panel B.ii: Email Sent by Mother cc'ing Mother Conditional On Calling				
Called Female | Call	0.35	0.12	0.08	0.21	0.48	0.83
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Called Male | Call	0.65	0.88	0.92	0.79	0.52	0.17
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Observations	3082	682	649	511	564	676

Notes: Standard errors are in parentheses. Observations do not have to be weighted in this table by whether the email sender is the mother or father in this table because the panels only show responses to emails from mother or father. Observations are weighted so that all message types have equal weighting.

Conditional on a call being made, sending the email from the father results in him being called 65% of the time (Figure 4 and Table 2, Panel B.ii, Column 1), meaning that external decision makers are still calling the mother one-third of the time even when she did not send the message. However, when the mother sends the message, 83% of the responding principals call her first (Figure 4 and Table 2, Panel A.ii, Column 1), resulting in the father being called less than one-fifth of the time. This highlights a ceiling that decision makers have on how much they will let a father be the primary contact for child-related tasks.

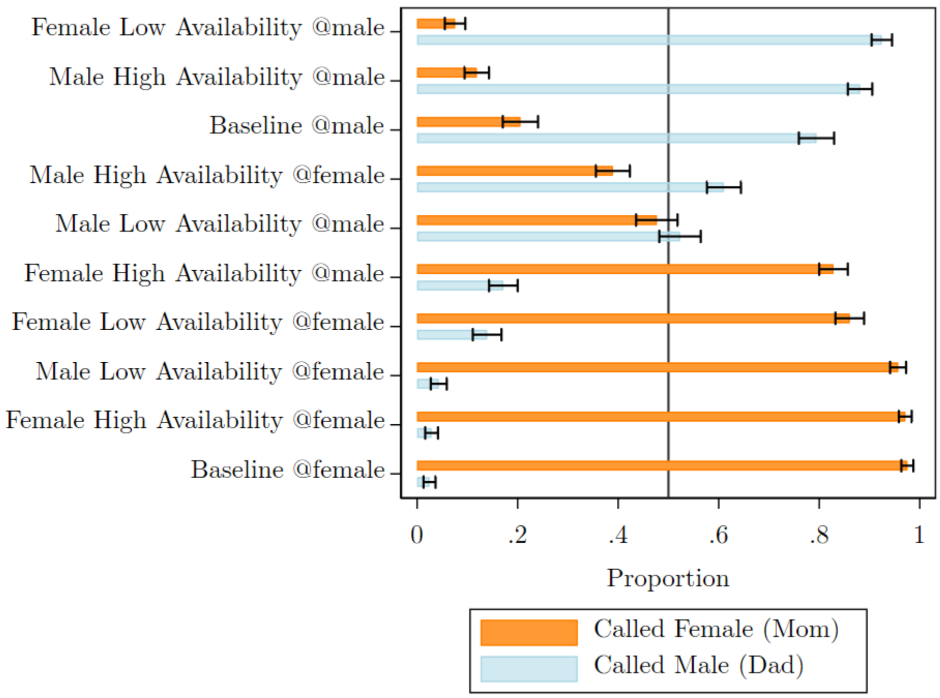
Examining the differences across treatment messages in more detail, we see that three of our messages, Baseline/LowMale/HighFemale, result in the mother being called over 95% of the time when she sends the email (bottom three rows of Figure 4). In contrast, none of our messages push the father to be called more than 95% of the time when he sends the email. This underscores a striking asymmetry in the effects of informational interventions on the gender gap in external demands for parental involvement and suggests that external decision makers have a ceiling on how much they will contact the father, while no such ceiling exists for mothers.

Last, something striking about Figure 4 is that almost none of our email treatment pairs result in a 50-50 split between calls to mothers and fathers despite many households reporting they would like closer to equal splits in parenting responsibilities. This may be because most schools, and other child-related activities, only allow two-parent households to denote a single "Primary Contact," essentially pushing the household toward a corner solution of always call mom or always call dad. This is likely an artifact of traditional gender norms where one parent focuses on housework, while the other focuses on work outside the home. The database systems that schools use push households toward a corner solution, when many households would prefer an interior solution.

5.3 Drivers of Gender Inequality

Our theoretical model described in Section 2 allows us to investigate whether the gender inequality we observe in the baseline message is driven by the decision maker's beliefs about parents' responsiveness or other deterrents. Intuitively, in the US, mothers are more likely to be stay-at-home parents than fathers (US Census Bureau, 2022). This general statistical information could lead decision makers to believe that female parents on average will be more responsive and as such will bias decision makers toward making more external demands of women. In Appendix K we show that these types of decision makers indeed report that they prefer to contact female parents because they believe mothers are more responsive but also

Figure 4: Outcomes in Main by Treatment and Email Sender



Notes: In this figure we show the proportion of decision makers choosing to call the female parent (mom) or the male parent (dad) conditional on a call being made by Treatment in our Main variation ($N = 6,382$) and whether the primary sender of the email was the female, "@female", or the male parent, "@male." Note, we always cc the other parent in all our emails. The ordering from top to bottom is sorted by "proportion called female-proportion called male". A vertical line is shown at the 50-50 equal split of calls between mothers and fathers, which is almost never where the actual observed data lays. Details of "No Call" are shown in Table 2. To obtain standard errors we regress dummy variables for our five messages interacted with the gender of the email sender on a binary variable for whether the female parent or male parent was called first.

because they believe mothers may be the primary contact about child-related topics, which we address in Section 5.3.2. Furthermore, in our own survey we find that female parents self-report being the first to respond 82% of the time, while male parents are the first to respond to the school only 42% of the time, indicating that decision makers correctly anticipate that mothers are more responsive. In Section 6.2 we discuss whether contacting the female parent first is in fact efficient under various definitions of efficiency.

Beyond responsiveness, there may be other deterrents affecting decision makers' choice to call a parent of a certain type. For example, they may prefer talking to mothers because they are more pleasant or prefer talking to fathers because they are better able to make decisions for the whole household in a patriarchal society. Alternatively, they may decide which parent to call based on the prevailing gender norms. There may also be other belief-based factors, unrelated to responsiveness. For example, in our specific setting, principals may believe that mothers are easier to convince to enroll in their school, which may explain why they are more likely to call mothers than fathers. Finally, institutional or systemic discrimination may also lead to the gender gaps that we observe. While we cannot disentangle the role of each possible factor in our experiment, we can shed light on the relative role of beliefs about responsiveness vis-a-vis other deterrents.

5.3.1 Parameter Estimates

We find that our parameter estimate for the responsiveness of female parents is $\bar{r}_f = -0.35$, which is less than the analogous parameter for male parents $\bar{r}_m = -0.25$, although the difference is not statistically significant ($Prob > chi2 = 0.60$ derived from results in Table A.2). We thus do not find support for Hypothesis 2. Although it's possible that we are missing an effect here due to imprecision, beliefs about responsiveness do not seem to be the main driver of the gender inequality in external demands for parents' time.

Next, we test if the gender inequality we document can be explained by other deterrents, as discussed by Hypothesis 3. We find that our parameter estimates for the residual term for male parents is greater than that for female parents; that is, $\bar{\delta}_m - \bar{\delta}_f = 0.48$ ($Prob > chi2 = 0.0008$). This is direct evidence that some of the gender inequality in demand for parents' involvement is driven by factors unrelated to beliefs about responsiveness. We investigate these factors below.

5.3.2 Beliefs about Parental Involvement and Expertise

It is likely that in deciding which parent to call, decision makers want to get a quick and useful response. Indeed, in our own survey, educators reported that they wanted to call mothers most often both because mothers were more responsive and because they were more likely to be the primary person making decisions about a wide range of child-related topics (e.g., sick child, child's allergies, school-related payments, volunteering at a book fair or at career day).

To better understand if our findings are partially driven by beliefs that child-related choices are primarily made by mothers, we added the following sentence to all our messages: "This is the type of decision we both want to be involved in equally." We sent out an additional 30,320 emails with this additional sentence that we call the equal decision variation. If beliefs that mothers primarily make child-related decisions are driving some of the inequality, then we would expect fewer calls to mothers with the addition of this sentence.

As detailed in Table 3, we find that mothers receive 11.7% of the calls and fathers 8.3% of the calls in the equal decision variation (see Appendix E for details by message variations). Conditional on a call being made, mothers get 58.7% of the calls in the equal decision variation, which is nearly identical to the results in the main variation. In that variation, 12.4% of the calls were made to mothers and 8.5% of the calls were made to fathers, but conditional on a call being made, 59.3% of the calls were made to the mother. Some of these differences are statistically significant but not economically significant. Overall, we see the ratio of calls to mothers versus fathers is almost exactly the same in the equal decision variation relative to the main variation, which makes it all the less likely that our findings are driven by beliefs that mothers primarily make child-related choices.

Table 3: **Summary Statistics by Variation (All Treatments Combined)**

	Panel A: All Outcomes			
	(1) Main	(2) Equal Decision	(3) Full Time	(4) Payments
Called Female	0.124	0.117	0.113	0.100
	(0.002)	(0.002)	(0.003)	(0.003)
Called Male	0.085	0.083	0.077	0.067
	(0.002)	(0.002)	(0.003)	(0.002)
No Call	0.791	0.800	0.810	0.833
	(0.002)	(0.002)	(0.004)	(0.004)
Observations	30471	30320	9472	9808
	Panel B: Conditional on Calling			
Called Female | Call	0.593	0.587	0.594	0.600
	(0.006)	(0.006)	(0.012)	(0.012)
Called Male | Call	0.407	0.413	0.406	0.400
	(0.006)	(0.006)	(0.012)	(0.012)
Observations	6382	6046	1817	1636

Notes: Standard errors are in parentheses. Observations are weighted so that there is 50% of emails from a female parent and 50% from a male parent, and so that all message types have equal weighting. Outcomes by message sent within these variations are available in Appendix E.

5.3.3 Beliefs about Stay-at-Home Mothers

In the US, mothers are significantly more likely to be a stay-at-home parent than fathers (US Census Bureau, 2022). To better understand if our findings are partially driven by beliefs about stay-at-home parents being more likely to be female, we added the following sentence to all our messages: "We both work full time." This sentence is meant to shut down the mechanism that the mother is a stay-at-home parent. We call this the full-time variation, and we sent this to an additional 9,472 principals (see Appendix E for details by message variations).

We would expect fewer calls to mothers in our full-time variation if beliefs that mothers were more likely to be a stay-at-home parent are driving the gender inequality. We do not find evidence of this as shown in Table 3. The rates of calls to mothers and fathers are quite similar in the full-time variation and the main variation. In the full-time variation mothers receive 11.3% of the calls and fathers receive 7.7% of the calls, which is almost identical to the main variation. Conditional on a call being made, the mother is called 59.4% of the time. In fact the ratio of calls to mothers versus fathers rises very slightly from 59.3% in the main variation when we include information that shuts down the idea that the mother is a stay-at-home parent.

5.3.4 Gender Norms

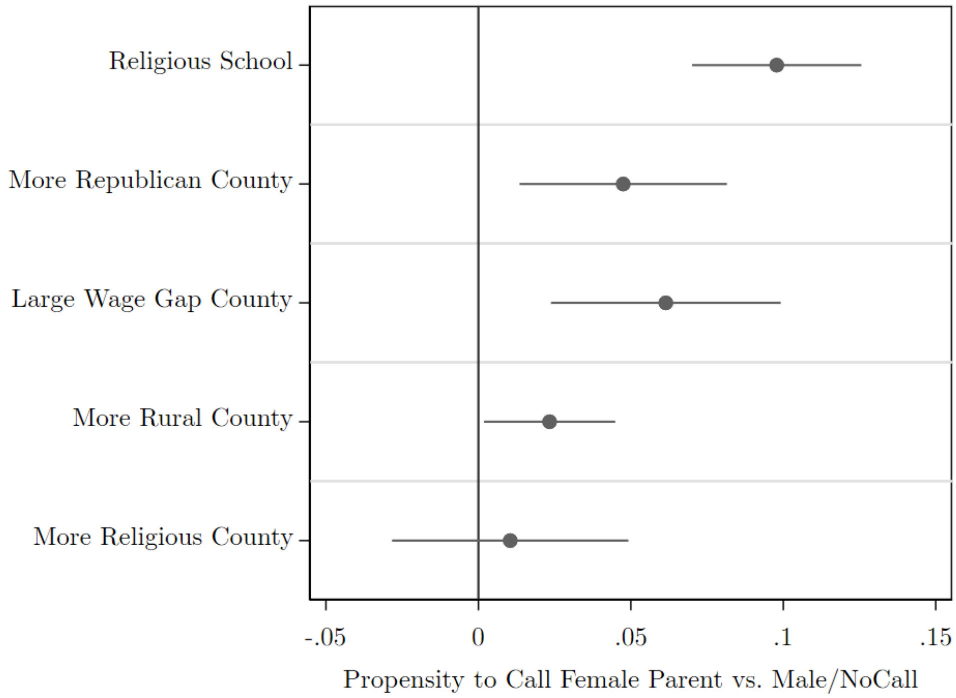
Another mechanism that could explain the gender gap in external demands for parental involvement that we document in our experiment is a strong gender norm governing interactions between decision makers and parents. As prior studies have shown, despite women's considerable advances in education and labor market outcomes in recent years, social norms about gender identity have persisted and still impact a wide range of economic and social outcomes for women, from labor force participation and earnings to marriage formation, fertility, and the division of home production (Bertrand et al., 2015; Kerwin et al., 2022; Jayachandran, 2021). While we do not have a precise measure of the gender norms of the principals or schools in our sample, we use multiple related measures to investigate whether gender norms may be driving some of the gender inequality in our setting.

Figure 5 shows that a variety of variables that might be associated with more traditional gender norms are also associated with a higher rate of decision makers calling the female parent in response to the baseline message in the main variation. At the most specific level, the school level, we observe whether a school is a religious school, which might denote that it believes in more traditional gender norms. If our results are in part driven by these gender norms, we would expect greater gender inequality in calls from religious than non-religious schools.[25] This is exactly what we find, especially in the unconditional call proportions. In particular, baseline unconditional call-back rates for religious schools are 21% to mothers and 11% to fathers, versus 12% and 8% for non-religious private and public schools (see Table A.3).

We also link our schools to other indicators of gender norms in the county the school is located in. Specifically, we look at the proportion of Republican voters in the 2016 presiden-

[25]Ethnicity is another dimension along which we might see variation in gender norms. Prior studies, however, have not found strong evidence of this (Wilcox, 1989; Kluegel and Smith, 1986). Also, we found little previous work to support the idea that gender norms might vary by decision-maker gender, and indeed find little difference in the patterns by the gender of the principal (Figure D.1).

Figure 5: Propensity to Call Female Parent by Gender Norm Proxies



Notes: In this figure we show the coefficients from regressions predicting whether a female parent is called versus a male parent or no call as a function of proxies for more traditional gender norms (religious school, republican county, large gender wage gap, more rural, more religious). The details of how these proxies are defined and more details are available in Table A.3. Observations are weighted so that 50% of emails come from a female parent and 50% from a male parent (always CCing the other parent).

tial election, the median wage gap between male and female workers, whether the county is more rural, and whether the county has a higher rate of religious attendance. We find that principals call the mother more in counties that have a high Republican vote share and greater gender wage gaps and are more rural and religious (see Figure 5 and Table A.3).[26] These findings provide strong evidence of the important role that gender norms play in perpetuating the gender inequality in external demands for parents' time.

5.4 Gender Inequality in More Male-Stereotyped Domains

It is possible that both male and female parents are fielding similar numbers of external requests but certain types of requests are associated with the female or male domain. Our

[26] Additionally, we can measure gender norms directly using a sexism index based on data from the General Social Survey (GSS) but these data are only available at the state level. Matching at the state level for an individual school/principal decision makes this measure quite noisy. For example, New York State has a very centrist sexism index, but this masks that New York City is likely relatively non-sexist, while upstate New York may be more sexist. Here we do not observe the same pattern of greater inequality in calls in more sexist states (Table A.3). We believe this is because measuring norms at the state level is too inexact.

own survey (Appendix K) found that within the school setting, educators stated they most heavily favored calling the mother for a child being sick, for volunteering at a book fair, and when dealing with allergies. While the educators still favored the mother, they did so to a lesser degree for requests to volunteer for a career day and to discuss school payments, and others have found that finances tend to be a more stereotypical male domain (Lin et al., 2022).

To test if fathers are contacted more often in more male-stereotyped domains, we fielded an additional variation of our email messages that stated "We are searching for schools for our child and are especially interested in discussing school fees and other expenses." In this variation, we observe fewer calls to mothers, with only 10.0% of principals calling the female parent when the sentence about fees is added (versus 12.4% in the main variation $p = 0.00$, Table 3). However, we also see fewer calls to fathers, with only 6.7% of principals calling fathers in response to the emails about fees (versus 8.5% $p = 0.00$). The actual rate of calling mothers versus fathers conditional on a call being made is not statistically significantly different from the main variation at 59.3% (versus 60.0%). Thus, even in a stereotypical male domain within the school setting, we do not see a shifting of the calls from mothers to fathers.

6 Discussion

6.1 Generalizability of the Results

Our theoretical model provides intuition for the underlying drivers behind the striking gender inequality in demand for parental involvement that we document in our field experiment. The framework can be adapted to derive results in other settings, where the forces at work could differ.

Setting. As a first step toward understanding gender inequality in external demands, we conduct a field experiment in the K-12 school setting. This setting is broadly applicable to the general population since 40% of households in the US have school-aged children. However, it is possible that our results could differ by setting, with some settings showing a similar skew toward calling mothers (e.g., doctors, dentists, daycare programs, extracurricular activities, summer camps) and others likely showing a skew toward calling fathers (e.g., household retirement planning, Boy Scouts, sports). Indeed our own survey of households indicates that many settings show a skew towards mothers as reflected in Figure 1.[27] Furthermore, because we are interested in gender inequality, we only study two-parent households with one male and one female parent, while other household arrangements exist and are important to study. We discuss these in Appendix H.

Lower Bound. We believe that the inequality we document is likely a lower bound on the total inequality that women face in external demands on their time versus men. It is likely that women experience more interruptions regarding the needs of not only their children but also of any adult family members who require caretaking (AARP, 2020). Also, researchers are increasingly finding that women shoulder a disproportionately large share of the cognitive load associated with managing a household (Daminger, 2019). Activities such as coordinating childcare, thinking about and anticipating future household needs, and other forms of invisible mental labor tend to be highly gendered and impose substantial disruptions to women's paid work.

Using the language of List (2020), this study represents a "first wave" study in which we focus on establishing causality and illuminating mechanisms with the help of a theoretical model. Although our evidence comes from a particular setting (schools) and a specific decision maker (school principal), our conceptual framework and research design can be adapted to other settings and to adjacent research questions.

6.2 Efficiency

Multiple parties are involved in the interaction that we investigate: the parents, the external decision maker (in our case the school), the child, and the parent's employers if employed. With multiple parties involved and many trade-offs to consider, it is not readily apparent what the most efficient allocation of calls between mothers and fathers is. We discuss this below.

Parents. Survey evidence indicates that mothers, on average, wish they were contacted less often about child-related needs than they currently are (see Appendix K.2). The existing skew toward mothers contributes to gender gaps in a wide range of labor market and educational outcomes, including career trajectory, occupational choice, and earnings. Workday disruptions stemming from child-related interruptions have also been linked to declines in

[27] While our field experiment employs a one-shot request to a school from a new parent, others have studied repeated interactions between schools and parents (see Johnson et al. (2023) for a review).

women's physical and mental health (Zamarro and Prados, 2021). Furthermore, contacting the person the household indicates has more availability would likely reduce parents' stress levels; such reductions in stress are associated with better parenting (Conger et al., 2010).

In our experimental data, even when the email comes from the father and he signals that he has high availability, 12% of the calls are still directed to mothers (Table 2). This indicates that households that want a more egalitarian division of child-related tasks and household labor, specifically fathers who want to be more involved, may be limited in achieving their goals in this area. Therefore, the current inequality in demands for parental involvement appears to be inefficient for parents.

Finally, even if we assume that men and women *on average* have different comparative advantages, there is a distribution of these skills within each gender. This implies that pairwise households differ from the population average, resulting in deadweight loss due to inefficiencies within households. This further suggests that reducing the restrictions placed on households by institutions would lead to a decrease in the deadweight loss.

External Decision Makers. Decision makers may have multiple competing objectives. In our model (Section 2), the decision maker in the short run is maximizing the likelihood of a useful response. However, in the long run, an entity (school, church, extracurricular program, doctor) may find it desirable to have a more diverse set of parents involved (e.g., not skewed toward mothers), and they may also prefer to have more parents (e.g., both parents versus one) involved (Clark et al., 1980). A less myopic decision maker may want to call the father even if they believe he is less likely to respond or may provide a less useful response. We believe work on these trade-offs is an important area for future research.

Child. The skew toward mothers being called more may be welfare harming for children given the extensive evidence that children benefit from having both fathers and mothers involved (Pleck, 2007; Nakata, 2023). Yet, research on the engagement of fathers in child-related social services has found that along with gendered and cultural factors that support preference for the mother, the institutional aspects of social services result in partial or full exclusions of fathers from child-related interventions (Perez-Vaisvidovsky et al., 2023). This implies considerable welfare costs for children.

Parents' Employers and Economic Efficiency. Parents' employers would like to minimize interruptions to their employees' workday. If the school is going to contact a parent, each employer would prefer that the school contacts the parent it does not employ. This has the flavor of a zero-sum game between the two employers. However, it would be most efficient, from the standpoint of both the mother's and the father's workplaces (and the overall economy), for the parent who has signaled more availability to be contacted provided that the household has information about which parent is a more productive worker. This would protect the more productive worker's time, increasing the combined output from the two parents. We find evidence that decision makers listen to these signals but do not fully integrate them, as 26% of the calls still go to mothers even when the father states he is highly available (Table 1).

Further investigation of the trade-offs each party faces, and how a social planner might weigh the needs of the various parties, is an important next step in this research agenda.

7 Conclusion

In this paper, we investigate a novel gender inequality in external demands for parental involvement. We develop a theoretical model that motivates the design of a large-scale field experiment in a K-12 school setting. In this experiment, we send emails to over 80,000 US school principals with a general inquiry about the school and a request to call one of the parents back. We randomly vary signals about parents' availability as well as which parent sends the email.

We document a striking gender inequality in responses. Conditional on receiving a callback, mothers are called first 40% more than fathers. To our knowledge, this provides the first empirical evidence of significant gender inequality in external demands for parental time. We show that signaling the availability of fathers mitigates this inequality and causes mothers to be called less than half the time. However, we observe a striking asymmetry in the effect of our informational interventions. Specifically, even when fathers strongly signal their availability, mothers are still called 26% of the time. In contrast, signals that reinforce stereotypes about mothers being more available cause them to receive 90% of the calls. Notably, even when the email comes from the father and he signals his availability, 12% of the calls are still directed to mothers. In contrast, fathers receive only 3% of the calls when mothers are the primary senders and signal that they are available. This underscores a ceiling on the degree to which informational signals can mitigate gender inequality in external demands for parental involvement.

Our theoretical model allows us to disentangle the mechanisms underlying any differential demand for parental time into beliefs about responsiveness versus other deterrents. We measure the impact of beliefs about responsiveness by randomizing the signals we send to decision makers about the availability and/or involvement of a specific parent, while the other factors are measured as a residual term in our model. We find that decision makers hold similar beliefs about the responsiveness of male and female parents in our setting. In contrast, we find that the inequality that we document is driven in part by differences in the residual. We test several potential deterrents, including beliefs about mothers being more likely to be a stay-at-home parent, beliefs that mothers are the primary decision maker on child-related choices, and the role of gender norms. We find evidence that gender norms are in part responsible for the gender gap in external demands for parental involvement.

We believe that the patterns we document represent a lower bound on the overall gender inequality in demands for parental involvement. The school setting is only one of many domains where gender differences in external demands on parents' time lead to disproportionate workday interruptions for mothers. While it is possible that fathers receive more requests in certain (male-stereotyped) domains, we do not observe this to be the case even in the most male-stereotypical task in our experiment (asking about school payments).

The gender gap that we document can have detrimental and persistent effects on women's career trajectories. More frequent workday interruptions for women versus men have been linked to a wide range of important economic outcomes, including occupational choice, human capital accumulation, and promotions. Furthermore, if women are disproportionately shouldering child-related and household tasks, they incur substantial personal costs, including to physical and mental health. Investigating the source of these inequalities and documenting that they are in part driven by external demands informs policies aimed at mitigating the gaps. As our findings indicate, both households' and external decision makers' actions can affect the size of the inequality. To mitigate this gap, it is essential for parents to signal the availability of fathers and for schools to foster more equitable parental involvement.

References

AARP, "Caregiving in the United States 2020," 2020.

Adams-Prassl, Abi, Kotaro Hara, Kristy Milland, and Chris Callison-Burch, "The gender wage gap on an online labor market: the cost of interruptions," *Review of Economics and Statistics*, 2023.

Agan, Amanda and Sonja Starr, "Ban the box, criminal records, and racial discrimination: A field experiment," *The Quarterly Journal of Economics*, 2018, *133* (1), 191–235.

Aguiar, Mark and Erik Hurst, "Measuring Trends in Leisure: The Allocation of Time Over Five Decades," *The Quarterly Journal of Economics*, 2007.

Ahmed, Ali, Mats Hammarstedt, and Karl Karlsson, "Do Swedish schools discriminate against children with disabilities?," *Working Paper*, 2020.

Aigner, Dennis J and Glen G Cain, "Statistical theories of discrimination in labor markets," *ILR Review*, 1977, *30* (2), 175–187.

Albanese, Andrea, Adrián Nieto, and Konstantinos Tatsiramos, "Job Location Decisions and the Effect of Children on the Employment Gender Gap," *CESifo Paper 9792*, 2022.

Amuedo-Dorantes, Catalina, Miriam Marcén, Marina Morales, and Almudena Sevilla, "COVID-19 School Closures and Parental Labor Supply in the United States," IZA Discussion Papers 13827, Institute of Labor Economics (IZA) October 2020.

Anderson, Deborah J, Melissa Binder, and Kate Krause, "The motherhood wage penalty: Which mothers pay it and why?," *American economic review*, 2002, *92* (2), 354–358.

Angelov, Nikolay, Per Johansson, and Erica Lindahl, "Parenthood and the gender gap in pay," *Journal of labor economics*, 2016, *34* (3), 545–579.

APA, "American psychological association," 2021.

Arrow, Kenneth, "The theory of discrimination,," in Orley Ashenfelter and Albert Rees, eds., *Discrimination in Labor Markets*, Princeton: Princeton University Press, 1973, pp. 3–3.

Becker, Gary S, *The economics of discrimination*, University of Chicago press, 1957.

Belkin, Lisa, "Dads in the PTA," *New York Times Mortherlode Blog*, 2009.

Bergman, Peter and Isaac McFarlin Jr, "Education for all? A nationwide audit study of school choice," *Working Paper*, 2018.

Bertrand, Marianne and Esther Duflo, "Field experiments on discrimination," *Handbook of economic field experiments*, 2017, *1*, 309–393.

_ **and Sendhil Mullainathan**, "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American economic review*, 2004, *94* (4), 991–1013.

_ **, Emir Kamenica, and Jessica Pan**, " Gender Identity and Relative Income within Households *," *The Quarterly Journal of Economics*, 01 2015, *130* (2), 571–614.

Bianchi, Suzanne M, John P Robinson, and Melissa A Milke, *The changing rhythms of American family life*, Russell Sage Foundation, 2006.

_ **, Liana C Sayer, Melissa A Milkie, and John P Robinson**, "Housework: Who did, does or will do it, and how much does it matter?," *Social forces*, 2012, *91* (1), 55–63.

BLS, "Bureau of Labor Statistics," 2021.

Bohren, J Aislinn, Alex Imas, and Michael Rosenberg, "The dynamics of discrimination: Theory and evidence," *American economic review*, 2019, *109* (10), 3395–3436.

—, Peter Hull, and Alex Imas, "Systemic discrimination: Theory and measurement," Technical Report, National Bureau of Economic Research 2022.

Boye, Katarina, "Can you stay home today? Parents' occupations, relative resources and division of care leave for sick children," *Acta Sociologica*, 2015, *58* (4), 357–370.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais, "'Acting wife': Marriage market incentives and labor market investments," *American Economic Review*, 2017, *107* (11), 3288–3319.

Cantet, Natalia, Brian Feld, and Monica Hernandez, "Is there discrimination against children of same-sex households?," *Working Paper*, 2022.

Charmes, Jacques, "The Unpaid Care Work and the Labour Market. An analysis of time use data based on the latest World Compilation of Time-use Surveys," *International Labour Office–Geneva: ILO*, 2019.

Charness, Gary, Anya Samek, and Jeroen van de Ven, "What is considered deception in experimental economics?," *Experimental Economics*, 2022, *25* (2), 385–412.

Ciasullo, Ludovica and Martina Uccioli, "What Works for Working Mothers? A Regular Schedule Lowers the Child Penalty," *Working Paper*, 2023.

Clark, David L, Linda S Lotto, and Martha M McCarthy, "Factors associated with success in urban elementary schools," *The Phi Delta Kappan*, 1980, *61* (7), 467–470.

Conger, Rand D, Katherine J Conger, and Monica J Martin, "Socioeconomic status, family processes, and individual development," *Journal of marriage and family*, 2010, *72* (3), 685–704.

Cortes, Patricia and Jessica Pan, "Children and the Remaining Gender Gaps in the Labor Market," *Working Paper*, 2021.

Couch, Kenneth A, Robert W Fairlie, and Huanan Xu, "The evolving impacts of the COVID-19 pandemic on gender inequality in the US labor market: The COVID motherhood penalty," *Economic Inquiry*, 2022, *60* (2), 485–507.

Council, National Research, "Measuring Racial Discrimination," 2004.

Cowan, Benjamin W, Todd R Jones, and Jeffrey M Swigert, "Parental and Student Time Use Around the Academic Year," Technical Report, National Bureau of Economic Research 2023.

Craig, Lyn and Killian Mullan, "How mothers and fathers share childcare: A cross-national time-use comparison," *American sociological review*, 2011, *76* (6), 834–861.

Cubas, G., C. Juhn, and P. Silos, "Coordinated Work Schedules and the Gender Wage Gap," *Economic Journal*, 2022.

Cubas, German, Chinhui Juhn, and Pedro Silos, "Work-Care Balance over the Day and the Gender Wage Gap," in "AEA Papers and Proceedings," Vol. 111 2021, pp. 149–53.

Daly, Moira and Fane Groes, "Who takes the child to the doctor? Mom, pretty much all of the time," *Applied Economics Letters*, 2017, *24* (17), 1267–1276.

Daminger, Allison, "The cognitive dimension of household labor," *American Sociological Review*, 2019, *84* (4), 609–633.

Diaz-Serrano, Luis and Enric Meix-Llop, "Do schools discriminate against homosexual parents? Evidence from a randomized correspondence experiment," *Economics of Education Review*, 2016, *53*, 133–142.

Duchini, Emma and Clémentine Van Effenterre, "School schedule and the gender pay gap," *Journal of Human Resources*, 2022.

Erosa, Andrés, Luisa Fuster, Gueorgui Kambourov, and Richard Rogerson, "Hours, Occupations, and Gender Differences in Labor Market Outcomes," *American Economic Journal: Macroeconomics*, 2022, *14* (3), 543–90.

Garcia, Kairon Shayne D and Benjamin W Cowan, "The impact of school and childcare closures on labor market outcomes during the COVID-19 pandemic," Technical Report, National Bureau of Economic Research 2022.

Gicheva, Dora, "Working long hours and early career outcomes in the high-end labor market," *Journal of Labor Economics*, 2013, *31* (4), 785–824.

Goldin, Claudia, "A grand gender convergence: Its last chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

Hansen, Benjamin, Joseph Sabia, and Jessamyn Schaller, "Schools, Job Flexibility, and Married Women's Labor Supply," *Working Paper*, 2022.

Heggeness, Misty L, "Estimating the immediate impact of the COVID-19 shock on parental attachment to the labor market and the double bind of mothers," *Review of Economics of the Household*, 2020, *18* (4), 1053–1078.

Hermes, Henning, Philipp Lergetporer, Philipp Peter, Fabian Mierisch, and Simon Wiederhold, "Males Should Mail? Gender Discrimination in Access to Child Care," in "AEA Papers and Proceedings" 2023.

Islam, Asad, Debayan Pakrashi, Liang C Wang, and Yves Zenou, "Determining the Extent of Statistical Discrimination: Evidence from a field experiment in India," 2018. Available at SSRN: https://ssrn.com/abstract=3185899.

Jayachandran, Seema, "Social Norms as a Barrier to Women's Employment in Developing Countries," *IMF Economic Review*, 2021.

Johnson, R., V. Mei, C. Hatzisamoura, and K. Gan, "Does the messenger matter? A randomized evaluation of the impact of providing teachers with technology for positive family engagement on student attendance," *The Lab@DC Working Paper*, 2023, *0* (0), 10443894221145751.

Kerwin, Charles, Jonathan Guryan, and Jessica Pan, "The Effects of Sexism on American Women: The Role of Norms vs. Discrimination," *Journal of Human Resources*, 2022.

Kleven, Henrik, "The Geography of Child Penalties and Gender Norms: Evidence from the United States," *Working Paper*, 2023.

_ , Camille Landais, and Jakob Egholt Søgaard, "Children and gender inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 2019, *11* (4), 181–209.

_ , _ , and _ , "Does biology drive child penalties? evidence from biological and adoptive families," *American Economic Review: Insights*, 2021, *3* (2), 183–98.

Kline, Patrick, Evan K Rose, and Christopher R Walters, "Systemic Discrimination Among Large U.S. Employers*," *The Quarterly Journal of Economics*, 06 2022, *137* (4), 1963–2036.

Kluegel, J. and E. Smith, "Beliefs About Inequality: Americans' Views of What Is and What Ought to Be.," 1986.

Kuziemko, Ilyana, Jessica Pan, Jenny Shen, and Ebonya Washington, "The mommy effect: Do women anticipate the employment effects of motherhood?," Technical Report, National Bureau of Economic Research 2018.

Laouénan, Morgane and Roland Rathelot, "Can information reduce ethnic discrimination? evidence from airbnb," *American Economic Journal: Applied Economics*, 2022, *14* (1), 107–32.

Lee, Yun-Suk and Linda J Waite, "Husbands' and wives' time spent on housework: A comparison of measures," *Journal of marriage and family*, 2005, *67* (2), 328–336.

Lin, Emily, Joel Slemrod, Evelyn Smith, and Alexander Yuskavage, "Who's on (the 1040) First? Determinants and Consequences of Spouses' Name Order on Joint Returns," *Working Paper*, 2022.

List, John A, "The nature and extent of discrimination in the marketplace: Evidence from the field," *The Quarterly Journal of Economics*, 2004, *119* (1), 49–89.

_ , "Non est disputandum de generalizability? A glimpse into the external validity trial," Technical Report, NBER 2020.

Mas, Alexandre and Amanda Pallais, "Valuing alternative work arrangements," *American Economic Review*, 2017, *107* (12), 3722–59.

_ and _ , "Alternative work arrangements," *Annual Review of Economics*, 2020, *12*, 631–658.

McFadden, Daniel, "The measurement of urban travel demand," *Journal of public economics*, 1974, *3* (4), 303–328.

Montes, Joshua, Christopher Smith, and Isabel Leigh, "Caregiving for children and parental labor force participation during the pandemic," 2021.

Nakata, Kazuko, "Income, Father's and Mother's Time Investment and Children's Achievement," *Working Paper*, 2023, *0* (0).

NCES, "National Center For Education Statistics," 2021.

Neal, Zachary, Jennifer Watling Neal, and Amelia Piteo, "Call me maybe: using incentives and follow-ups to increase principals' survey response rates," *Journal of Research on Educational Effectiveness*, 2020, *13* (4), 784–793.

Oberfield, Zachary W and Matthew B Incantalupo, "Racial Discrimination and Street-Level Managers: Performance, Publicness, and Group Bias," *Public Administration Review*, 2021, *81* (6), 1055–1070.

Offer, Shira, "The costs of thinking about work and family: Mental labor, work–family spillover, and gender inequality among parents in dual-earner families," in "Sociological Forum," Vol. 29 Wiley Online Library 2014, pp. 916–936.

Oreopoulos, Philip, "Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes," *American Economic Journal: Economic Policy*, 2011, *3* (4), 148–71.

Perez-Vaisvidovsky, Nadav, Ayana Halpern, Reli Mizrahi, and Zhara Atalla, "Fathers Are Very Important, but They Aren't Our Contact Persons: The Primary Contact Person Assumption and the Absence of Fathers in Social Work Interventions," *Families in Society*, 2023, *0* (0), 10443894221145751.

Pertold-Gebicka, Barbara, Filip Pertold, and Nabanita Datta Gupta, "Employment adjustments around childbirth," *IZA Discussion Paper*, 2016.

Pew Research Center, "Raising Kids and Running a Household: How Working Parents Share the Load," 2015.

Phelps, Edmund S, "The statistical theory of racism and sexism," *The american economic review*, 1972, *62* (4), 659–661.

Pleck, Joseph H, "Why could father involvement benefit children? Theoretical perspectives," *Applied development science*, 2007, *11* (4), 196–202.

Powell, Walter W and Paul J DiMaggio, *The new institutionalism in organizational analysis*, University of Chicago press, 2012.

Price, Brendan M and Melanie Wasserman, "The summer drop in female employment," *Working Paper*, 2022.

Russell, Lauren and Chuxuan Sun, "The Effect of Mandatory Child Care Center Closures on Women's Labor Market Outcomes During the COVID-19 Pandemic," 2020.

Schoonbroodt, Alice, "Parental child care during and outside of typical work hours," *Review of Economics of the Household*, 2018, *16* (2), 453–476.

Scott, W Richard, *Institutions and organizations: Ideas, interests, and identities*, Sage publications, 2013.

Scottland, Fathers' Network, "Why Fathers' Involvement Matters," 2020.

Sevilla, Almudena and Sarah Smith, "Baby steps: The gender division of childcare during the COVID-19 pandemic," *Oxford Review of Economic Policy*, 2020, *36* (Supplement_1), S169–S186.

Small, Mario L and Devah Pager, "Sociological perspectives on racial discrimination," *Journal of Economic Perspectives*, 2020, *34* (2), 49–67.

Speer, Jamin and Fulya Ersoy, "Opening the Black Box of College Major Choice: Evidence from an Information Intervention," *Working Paper*, 2022, *22:666.*

Tzioumis, Konstantinos, "Demographic aspects of first names," *Scientific data*, 2018, *5* (1), 1–9.

US Census Bureau, "Income and Poverty in the United States: 2019," 2020.

__ , "Number of stay-at-home parents among opposite-sex married couple families with children under 15," 2022.

Wasserman, Melanie, "Hours constraints, occupational choice, and gender: Evidence from medical residents," *Review of Economic Studies*, 2022.

Wikle, Jocelyn and Clara Cullen, "The Developmental Course of Parental Time Investments in Children from Infancy to Late Adolescence," *Social Sciences*, 2023, *12* (2), 92.

Wilcox, C., "Race, gender role attitudes and support for feminism.," *Politics Quarterly*, 1989.

Zamarro, G and M. Prados, "Gender differences in couples' division of childcare, work and mental health during COVID-19.," *Review of Economics of the Household*, 2021, *19* (1), 11–40.

Appendix: For Online Publication Only

A Appendix Tables

Table A.1: **Multinomial Logit Models of Effect of Treatments on No Call, Call Male, or Call Female**

	(1)	(2)	(3)	(4)	(5)	(6)
No_Call						
High Male (Hm)	-0.62***	-0.66***	0.81***	0.85***	0.00	0.00
	(0.06)	(0.06)	(0.07)	(0.07)	(.)	(.)
Low Female (Lf)	-0.26***	-0.27***	0.23***	0.24***	0.00	0.00
	(0.07)	(0.07)	(0.06)	(0.06)	(.)	(.)
Low Male (Lm)	0.38***	0.39***	-0.22***	-0.23***	0.00	0.00
	(0.08)	(0.08)	(0.05)	(0.05)	(.)	(.)
High Female (Hf)	1.31***	1.35***	-0.48***	-0.51***	0.00	0.00
	(0.10)	(0.10)	(0.05)	(0.05)	(.)	(.)
Female_Call						
High Male (Hm)	-1.44***	-1.51***	0.00	0.00	-0.81***	-0.85***
	(0.08)	(0.09)	(.)	(.)	(0.07)	(0.07)
Low Female (Lf)	-0.49***	-0.51***	0.00	0.00	-0.23***	-0.24***
	(0.08)	(0.08)	(.)	(.)	(0.06)	(0.06)
Low Male (Lm)	0.59***	0.62***	0.00	0.00	0.22***	0.23***
	(0.09)	(0.09)	(.)	(.)	(0.05)	(0.05)
High Female (Hf)	1.79***	1.86***	0.00	0.00	0.48***	0.51***
	(0.11)	(0.11)	(.)	(.)	(0.05)	(0.05)
Male_Call						
High Male (Hm)	0.00	0.00	1.44***	1.51***	0.62***	0.66***
	(.)	(.)	(0.08)	(0.09)	(0.06)	(0.06)
Low Female (Lf)	0.00	0.00	0.49***	0.51***	0.26***	0.27***
	(.)	(.)	(0.08)	(0.08)	(0.07)	(0.07)
Low Male (Lm)	0.00	0.00	-0.59***	-0.62***	-0.38***	-0.39***
	(.)	(.)	(0.09)	(0.09)	(0.08)	(0.08)
High Female (Hf)	0.00	0.00	-1.79***	-1.86***	-1.31***	-1.35***
	(.)	(.)	(0.11)	(0.11)	(0.10)	(0.10)
Control Variables		Yes		Yes		Yes
Observations	30471	30471	30471	30471	30471	30471

Notes: This table presents the results of a multinomial logit model using a model like the one in Equation 10. The outcome variable takes three values: no call, call female, or call male. In this table we present the results with a base case of no call in columns (1) and (2), female call in columns (3) and (4), and male call in columns (5) and (6). The results from the three base cases are analogous and all three are presented to make specific comparisons more simple. Observations are weighted so that 50% of emails come from a female parent and 50% from a male parent. The outcomes with no controls from this table are represented visually in Figure B.1.

Table A.2: **Multinomial Logit Models For Theory Model**

	(1) Main	(2) Expertise	(3) Payment	(4) Full Time
Female_Call				
any_msg_M	-0.30***	-0.19***	-0.37***	-0.28**
	(0.05)	(0.05)	(0.10)	(0.10)
x_M	-0.51***	-0.36***	-0.59***	-0.43***
	(0.03)	(0.03)	(0.07)	(0.07)
any_msg_F	0.12*	0.31***	0.04	0.36***
	(0.05)	(0.05)	(0.10)	(0.09)
x_F	0.36***	0.36***	0.35***	0.27***
	(0.03)	(0.03)	(0.06)	(0.05)
Constant	-1.87***	-2.04***	-2.09***	-2.08***
	(0.04)	(0.04)	(0.08)	(0.08)
Male_Call				
any_msg_M	0.12*	0.20***	0.29*	0.28**
	(0.06)	(0.06)	(0.12)	(0.10)
x_M	0.50***	0.41***	0.55***	0.41***
	(0.03)	(0.03)	(0.06)	(0.06)
any_msg_F	-0.53***	-0.34***	-0.40**	-0.55***
	(0.07)	(0.06)	(0.15)	(0.12)
x_F	-0.78***	-0.49***	-0.80***	-0.74***
	(0.05)	(0.04)	(0.11)	(0.09)
Constant	-2.25***	-2.32***	-2.69***	-2.42***
	(0.05)	(0.05)	(0.10)	(0.08)
N	30471.00	30320.00	9808.00	9472.00

Notes: This table presents the results of a multinomial logit model using a model like the one in Equation 10. The outcome variable takes three values: no call, call female, or call male. The right-hand side variables are any_msg_M which takes the value 1 if a message was sent with a signal about the male parent (MaleHigh, MaleLow), while any_msg_F takes the value 1 if a message with a signal about the female parent was sent (FemaleHigh, FemaleLow). The variable x_M takes the value 1 if the MaleHigh message was sent, and -1 if the MaleLow message, 0 otherwise; x_F is defined analogously for messages about female parents. The right-hand side variables are discussed in Section 2. In this table we present the results with a base case of no call. Observations are weighted so that 50% of emails come from a female parent and 50% from a male parent (always CCing the other parent).

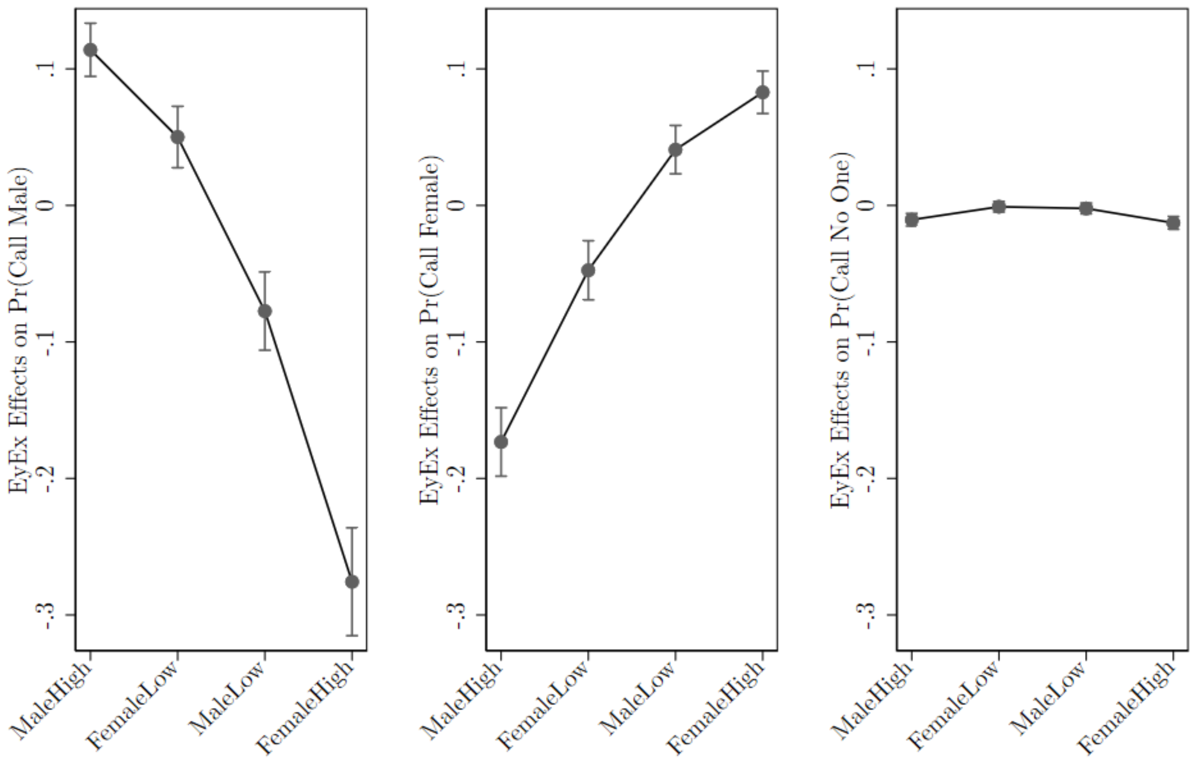
Table A.3: **More vs. Less Traditional Gender Norms Summary Statistics Baseline Message in Main Variation**

	(1) Non Religious School	(2) Religious School	(3) Low Repub. County	(4) High Repub. County	(5) Small Wage Gap County	(6) Large Wage Gap County	(7) Less Rural County	(8) More Rural County	(9) Less Religious County	(10) More Religious County	(11) Less Sexist State	(12) More Sexist State
Called Female	0.11	0.21	0.08	0.13	0.09	0.15	0.12	0.14	0.12	0.14	0.13	0.12
Called Male	0.08	0.12	0.07	0.09	0.04	0.12	0.08	0.09	0.09	0.06	0.08	0.07
No Call	0.80	0.67	0.85	0.78	0.87	0.74	0.80	0.77	0.78	0.80	0.79	0.81
Called Female | Call	0.58	0.63	0.52	0.58	0.69	0.56	0.59	0.62	0.58	0.69	0.63	0.61
Called Male | Call	0.42	0.37	0.48	0.42	0.31	0.44	0.41	0.38	0.42	0.31	0.37	0.39
Observations	4755	528	635	580	529	593	4439	1161	606	553	485	607

Notes: Religious school means the school is identified by our schools database as a religious school, while Non-Religious schools include public schools (non-charter) and private schools (non-religious). Low Republican means the school is located in a county at the 10th percentile or below of Republican vote share in the 2016 presidential election, while High Republican is at the 90th percentile or above. Small Wage Gap means the school is located in a county at the 10th percentile or below of the ratio between male-female median wages, while Large Wage Gap is at the 90th percentile or above. More Rural county means fewer than 250,000 population, while Less Rural is above that. Less Religious county is a county at the 10th percentile or lower for religious adherence, while More Religious county is above the 90th percentile as measure by the Association of Statisticians of American Religious Bodies (https://www.thearda.com/us-religion/sources-for-religious-congregations-membership-data#QR). Less Sexist State means the school is located in a state at the 10th percentile or below of the sexism index created by questions from the General Social Survey, while High Sexist State is at the 90th percentile or above. Observations are weighted so that 50% of emails come from a female parent and 50% from a male parent (always CCing the other parent).

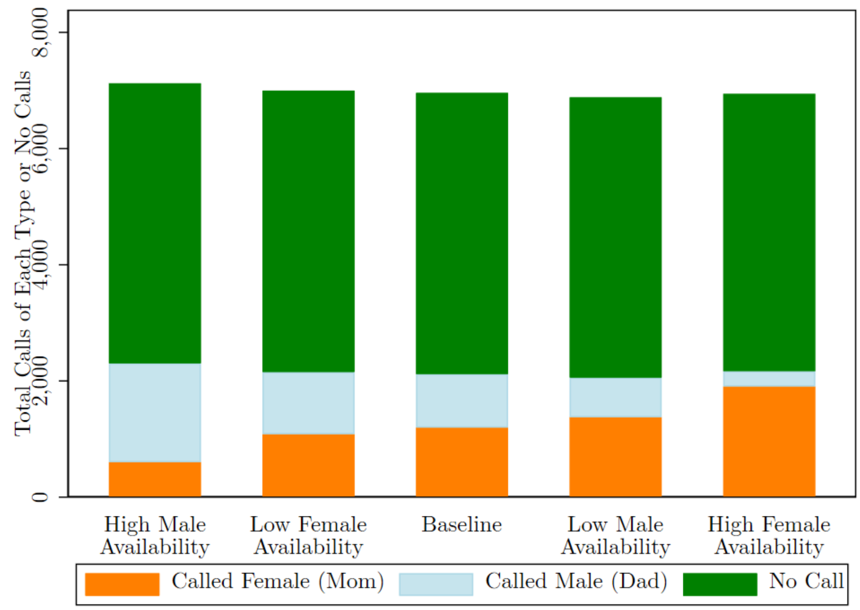
B Appendix Figures

Figure B.1: Effects by Treatment

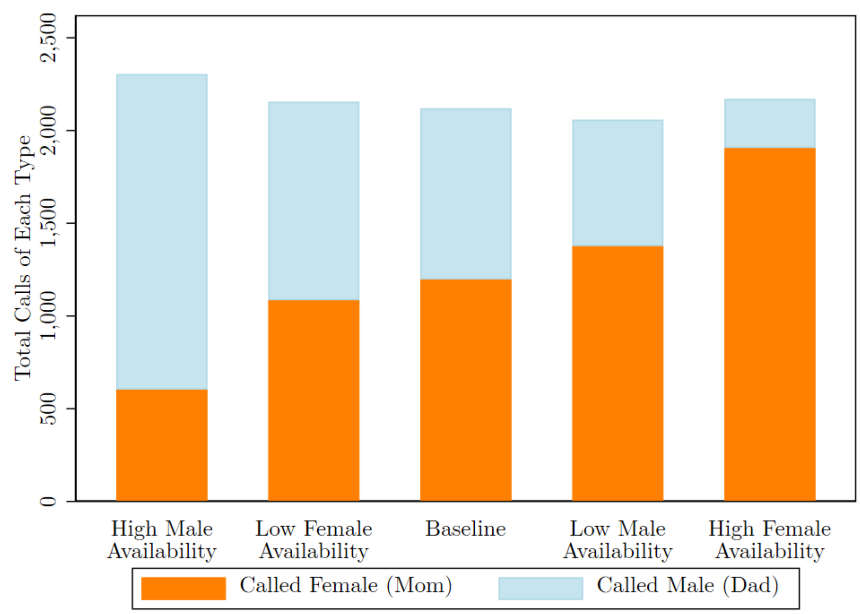


Notes: In this figure we show the results from a mutltinomial logit model using a model like Equation 10 which is detailed fully in Table A.1. This figure shows the marginal effects elasticities. Observations are weighted so that 50% of emails come from a female parent and 50% from a male parent (always CCing the other parent).

Figure B.2: Outcomes by Treatment in Main Variation for Multiple Calls



(a) All Outcomes



(b) Outcomes Conditional on Calling

Notes: In this figure we show the total number of no calls, calls the female parent (mom) or calls to the male parent (dad) by the message sent to the decision maker in our Main variation (see Figure 3 for proportions by only the first call or no call). Panel (a) represents three outcomes from 30, 471 decision makers, while panel (b) shows only the choices of those who made a phone call to at least one parent ($N = 7{,}778$). If decision makers were randomizing which parent they called we would expect the same proportion of calls to male and female parents. Two-way t-tests comparing No Call, Call Female, and Call Male are all statistically significant at the 5% level or below. Observations are weighted so that 50% of emails come from a female parent and 50% from a male parent (always CCing the other parent).

C Balance Tables

See Tables E.1, E.2, and E.3 for balance in the other Variations of our experiment. See Table L.6 for balance on observables when we do not re-weight to account for imbalance in emails sent from male versus female parents.

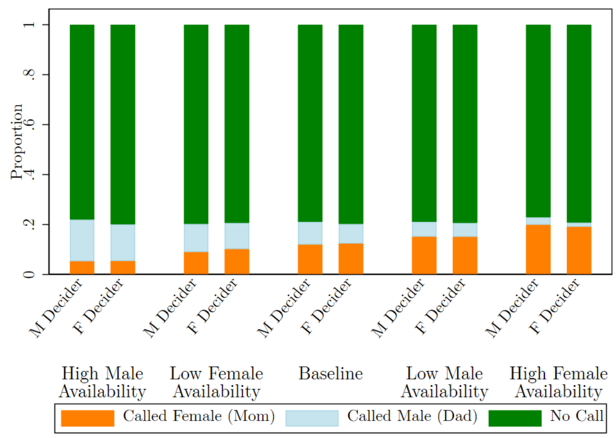
Table C.1: **Balance on Observable Attributes of Schools/Decision Makers by Treatment in Main Variation**

	(1) High Male (Hm)	(2) Low Female (Lf)	(3) Baseline (b)	(4) Low Male (Lm)	(5) High Female (Hf)
Elementary	0.48	0.49	0.51	0.50	0.50
Middle	0.14	0.14	0.14	0.15	0.15
High	0.19	0.20	0.20	0.19	0.20
Decison-Maker Female	0.57	0.58	0.59	0.59	0.58
PublicCharter	0.06	0.05	0.06	0.06	0.06
PublicNOTCharter	0.76	0.79	0.81	0.79	0.80
Private	0.18	0.16	0.13	0.15	0.14
FreeLunch	0.55	0.56	0.54	0.55	0.52
White	0.52	0.52	0.52	0.53	0.52
Black	0.14	0.15	0.14	0.14	0.15
Hispanic	0.23	0.23	0.23	0.23	0.23
FemaleEmail	0.50	0.50	0.50	0.50	0.50
Observations	7075	5931	5612	5700	6153

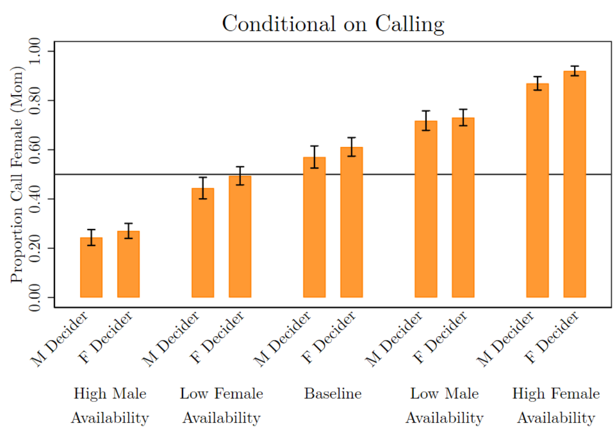
Notes: There is a small proportion of schools which are not elementary, middle or high schools (e.g. K–12 or preschools). The following variables are known only for non-private schools: FreeLunch, White, Black, Hispanic. DMFemale is whether the decision maker (the principal) has a first name that is female. Observations are weighted so that 50% of emails are from a female parent and 50% from a male parent.

D By Decision Maker Gender

Figure D.1: Outcomes By Principal Gender in "Main" Variation



(a) All Outcomes



(b) Outcomes Conditional On Calling

Notes: In this figure we show the differences between Female and Male principals in our "Main" variation. We predict principal gender based on their name. In panel (a) we show the proportion of decision makers choosing to make no call, call the female parent (mom) or the male parent (dad) by the message sent to the decision maker in our Main Variation. "M Decider" denotes a male principal and "F Decider" denotes a female principal. Panel (a) represents three outcomes from 30,471 decision makers in Main, while panel (b) shows only the choices of those who made a phone call to at least one parent ($N = 7,778$ in Main). In Panel B we regress dummy variables for our five messages on a binary variable for whether the female parent was called first or the male parent. If decision makers were randomizing which parent they called we would expect the same proportion of calls to male and female parents.

E Variations On Main Messages

E.1 Balance Tables For Variations

Table E.1: **Balance on Observable Attributes of Schools/Decision Makers By Treatment In Equal Decision Variation**

	(1) High Male (Hm)	(2) Low Female (Lf)	(3) Baseline (b)	(4) Low Male (Lm)	(5) High Female (Hf)
Elementary	0.50	0.50	0.49	0.48	0.48
Middle	0.15	0.15	0.13	0.14	0.14
High	0.20	0.20	0.18	0.19	0.18
Decison-Maker Female	0.58	0.58	0.57	0.58	0.57
PublicCharter	0.06	0.05	0.06	0.06	0.05
PublicNOTCharter	0.80	0.80	0.77	0.76	0.76
Private	0.14	0.14	0.18	0.18	0.18
FreeLunch	0.55	0.52	0.55	0.55	0.57
White	0.52	0.53	0.52	0.52	0.52
Black	0.15	0.15	0.15	0.15	0.15
Hispanic	0.23	0.23	0.23	0.23	0.23
FemaleEmail	0.50	0.50	0.50	0.50	0.50
Observations	5170	5558	6569	6755	6268

Notes: There is a small proportion of schools which are not Elementary, Middle or High Schools (e.g. K-12 or pre-schools). The following variables are only known for non-private schools: FreeLunch, White, Black, Hispanic. DMFemale is whether the decision maker (the principal) has a first name that is female. Observations are weighted so that there is 50% of emails from a female parent and 50% from a male parent.

Table E.2: **Balance on Observable Attributes of Schools/Decision Makers By Treatment In Full Time Variation**

	(1) High Male (Hm)	(2) Low Female (Lf)	(3) Baseline (b)	(4) Low Male (Lm)	(5) High Female (Hf)
Elementary	0.49	0.51	0.48	0.52	0.49
Middle	0.17	0.17	0.14	0.16	0.14
High	0.21	0.22	0.18	0.21	0.20
Decison-Maker Female	0.56	0.59	0.57	0.60	0.59
PublicCharter	0.06	0.06	0.05	0.06	0.05
PublicNOTCharter	0.80	0.82	0.73	0.81	0.77
Private	0.14	0.12	0.22	0.13	0.18
FreeLunch	0.55	0.56	0.53	0.55	0.54
White	0.52	0.52	0.52	0.53	0.52
Black	0.15	0.15	0.14	0.15	0.14
Hispanic	0.23	0.23	0.24	0.22	0.24
FemaleEmail	0.50	0.50	0.50	0.50	0.50
Observations	1785	1478	1943	1776	2490

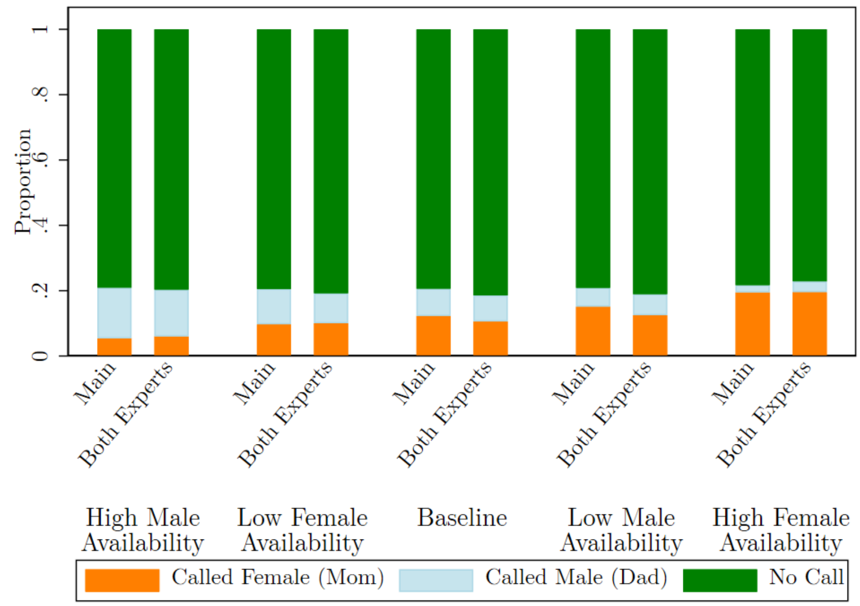
Notes: There is a small proportion of schools which are not Elementary, Middle or High Schools (e.g. K-12 or pre-schools). The following variables are only known for non-private schools: FreeLunch, White, Black, Hispanic. DMFemale is whether the decision maker (the principal) has a first name that is female. Observations are weighted so that there is 50% of emails from a female parent and 50% from a male parent.

Table E.3: **Balance on Observable Attributes of Schools/Decision Makers By Treatment In Payments Variation**

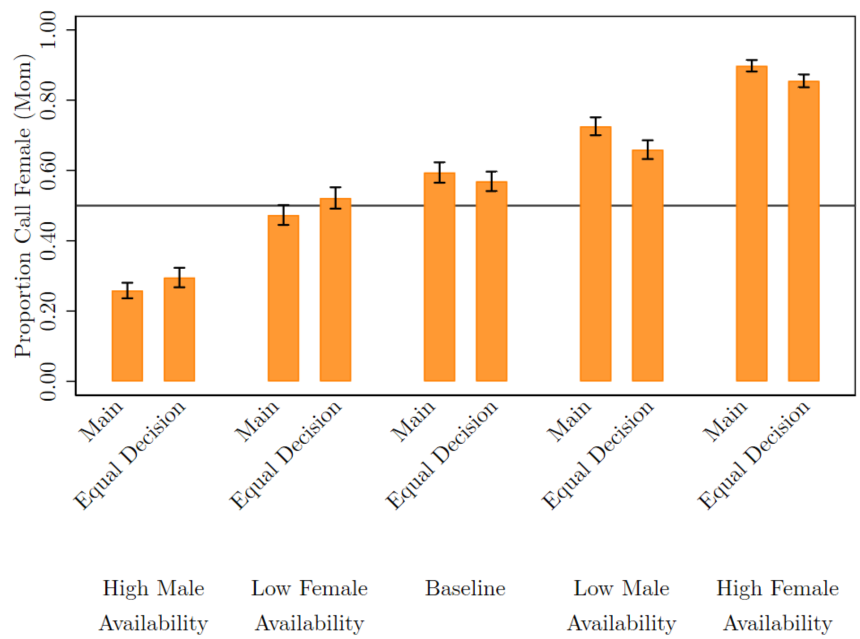
	(1) High Male (Hm)	(2) Low Female (Lf)	(3) Baseline (b)	(4) Low Male (Lm)	(5) High Female (Hf)
Elementary	0.50	0.50	0.50	0.49	0.52
Middle	0.15	0.14	0.16	0.15	0.17
High	0.19	0.19	0.22	0.21	0.20
Decison-Maker Female	0.58	0.60	0.58	0.58	0.58
PublicCharter	0.06	0.07	0.05	0.06	0.06
PublicNOTCharter	0.78	0.75	0.81	0.78	0.81
Private	0.17	0.18	0.14	0.16	0.12
FreeLunch	0.54	0.58	0.56	0.55	0.53
White	0.52	0.51	0.51	0.50	0.53
Black	0.15	0.15	0.15	0.15	0.15
Hispanic	0.23	0.23	0.23	0.25	0.22
FemaleEmail	0.50	0.50	0.50	0.50	0.50
Observations	2101	2153	1795	2333	1426

Notes: There is a small proportion of schools which are not Elementary, Middle or High Schools (e.g. K-12 or pre-schools). The following variables are only known for non-private schools: FreeLunch, White, Black, Hispanic. DMFemale is whether the decision maker (the principal) has a first name that is female. Observations are weighted so that there is 50% of emails from a female parent and 50% from a male parent.

Figure E.1: Outcomes By Treatment "Main" vs. "Equal Decision" Variations



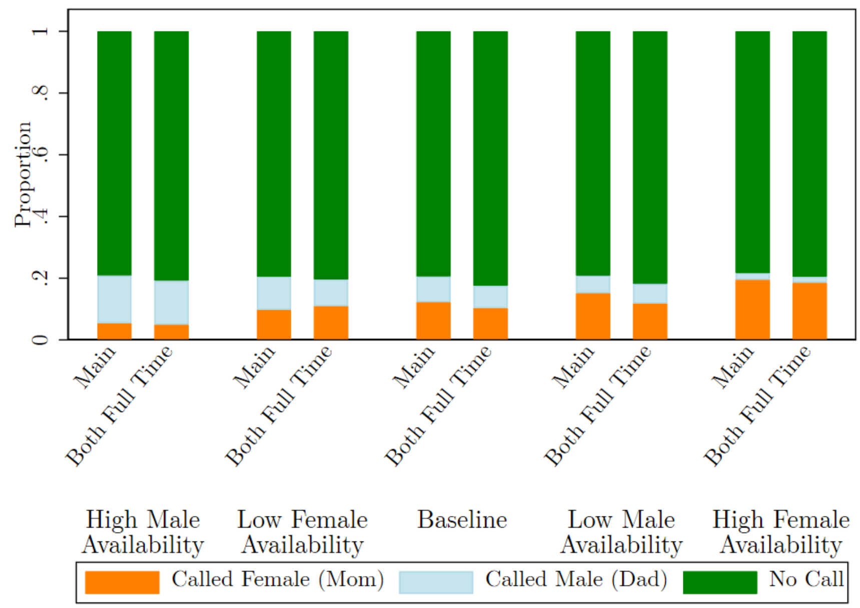
(a) All Outcomes



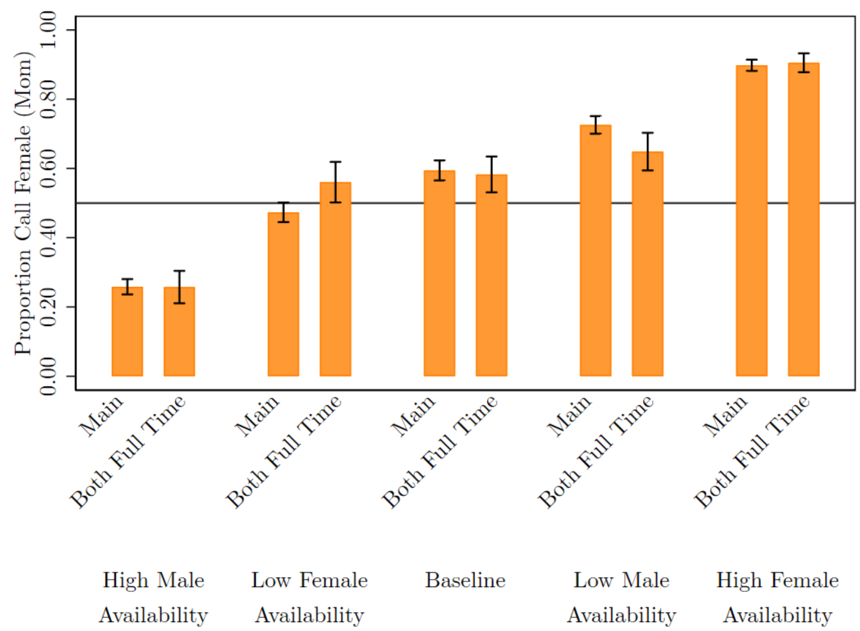
(b) Outcomes Conditional On Calling

Notes: In this figure we show the differences between our "Main" version of our emails and ones that have the addition of a sentence that states "This is the type of decision we both want to be involved in equally." In panel (a) we show the proportion of decision makers choosing to make no call, call the female parent (mom) or the male parent (dad) by the message sent to the decision maker in our Main Variation. Panel (a) represents three outcomes from $30,471$ decision makers in Main and $30,320$ in Equal Decision, while panel (b) shows only the choices of those who made a phone call to at least one parent ($N = 7,778$ in Main and $7,209$ in Equal Decision). In Panel B we regress dummy variables for our five messages on a binary variable for whether the female parent was called first or the male parent. If decision makers were randomizing which parent they called we would expect the same proportion of calls to male and female parents.

Figure E.2: Outcomes By Treatment "Main" vs. "Full Time" Variations



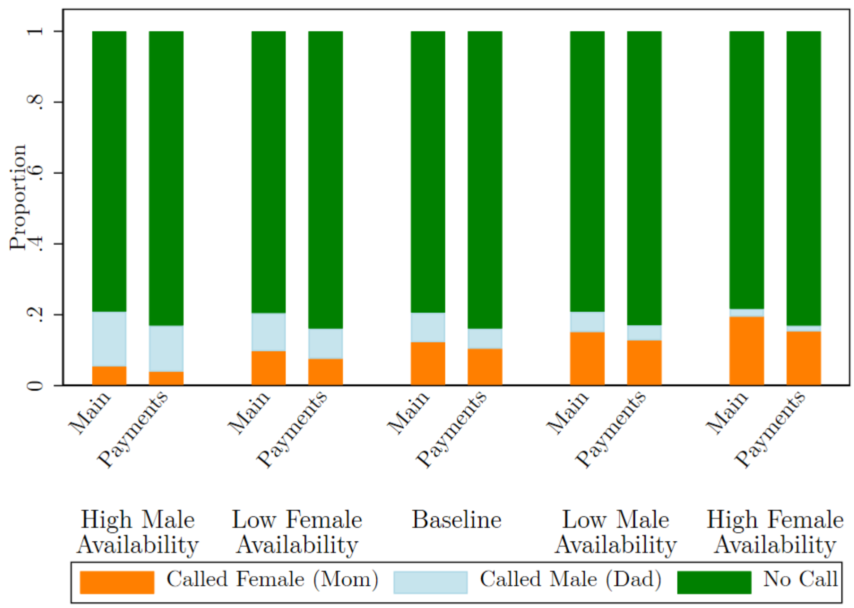
(a) All Outcomes



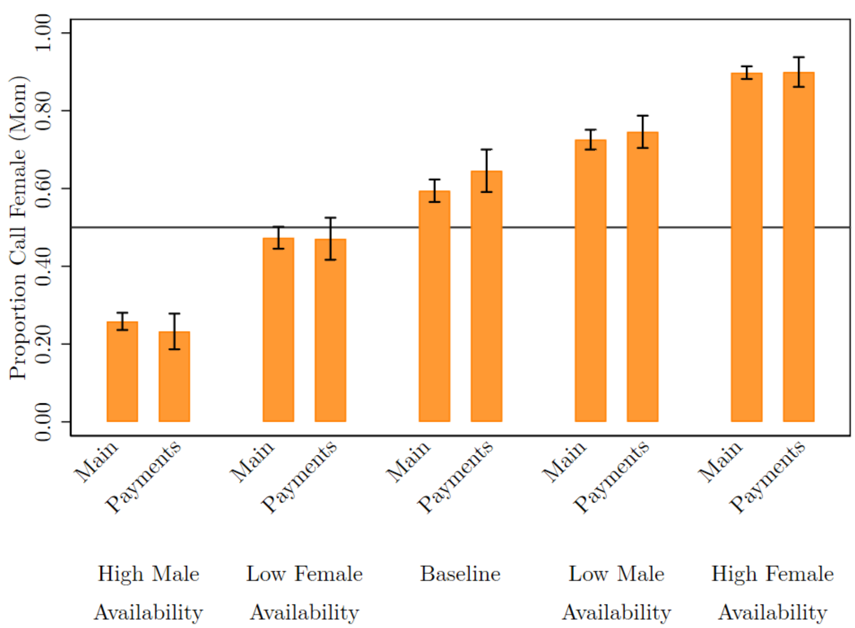
(b) Outcomes Conditional On Calling

Notes: In this figure we show the differences between our "Main" version of our emails and ones that have the addition of a sentence that states "We both work full-time." In panel (a) we show the proportion of decision makers choosing to make no call, call the female parent (mom) or the male parent (dad) by the message sent to the decision maker in our Main Variation. Panel (a) represents three outcomes from 30,471 decision makers in Main and 9,472 in Full Time, while panel (b) shows only the choices of those who made a phone call to at least one parent ($N = 7{,}778$ in Main and 2,175 in Full Time). In Panel B we regress dummy variables for our five messages on a binary variable for whether the female parent was called first or the male parent. If decision makers were randomizing which parent they called we would expect the same proportion of calls to male and female parents.

Figure E.3: Outcomes By Treatment "Main" vs. "Payments" Variations



(a) All Outcomes

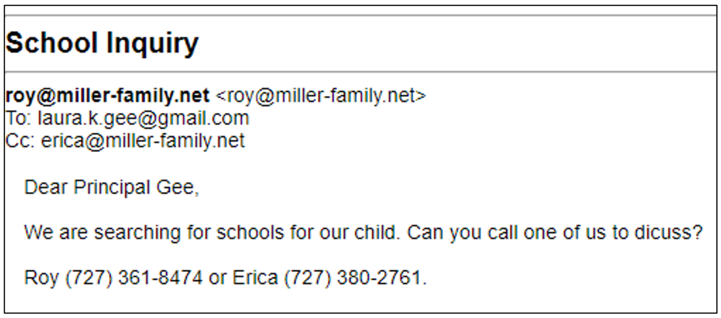


(b) Outcomes Conditional On Calling

Notes: In this figure we show the differences between our "Main" version of our emails and ones that have the addition of a clauses that states they are "especially interested in discussing school fees and other expenses." In panel (a) we show the proportion of decision makers choosing to make no call, call the female parent (mom) or the male parent (dad) by the message sent to the decision maker in our Main Variation. Panel (a) represents three outcomes from 30,471 decision makers in Main and 9,808 in Full Time, while panel (b) shows only the choices of those who made a phone call to at least one parent ($N = 7,778$ in Main and 9,472 in Full Time). In Panel B we regress dummy variables for our five messages on a binary variable for whether the female parent was called first or the male parent. If decision makers were randomizing which parent they called we would expect the same proportion of calls to male and female parents.

F Example Emails Full Text

Figure F.1: Main Baseline: no signal



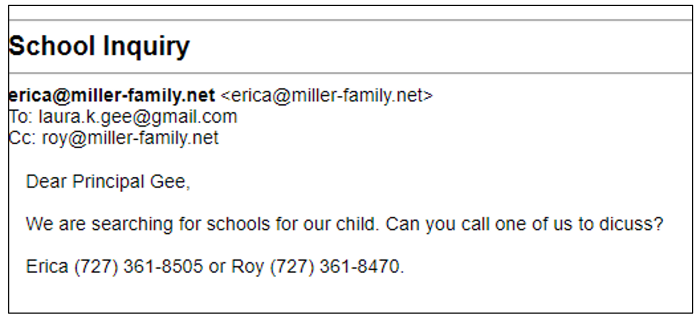
School Inquiry

roy@miller-family.net <roy@miller-family.net>
To: laura.k.gee@gmail.com
Cc: erica@miller-family.net

Dear Principal Gee,

We are searching for schools for our child. Can you call one of us to dicuss?

Roy (727) 361-8474 or Erica (727) 380-2761.



School Inquiry

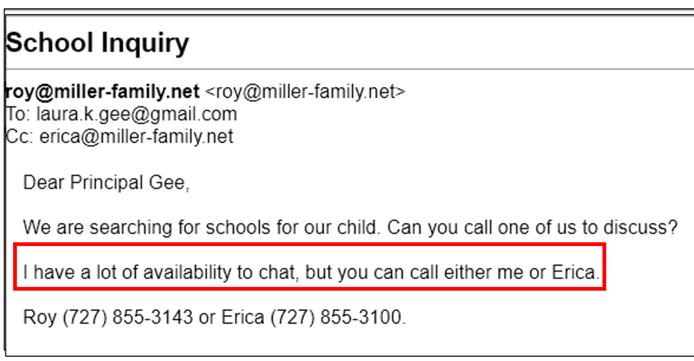
erica@miller-family.net <erica@miller-family.net>
To: laura.k.gee@gmail.com
Cc: roy@miller-family.net

Dear Principal Gee,

We are searching for schools for our child. Can you call one of us to dicuss?

Erica (727) 361-8505 or Roy (727) 361-8470.

Figure F.2: Main: High Male and Low Male Signal



School Inquiry

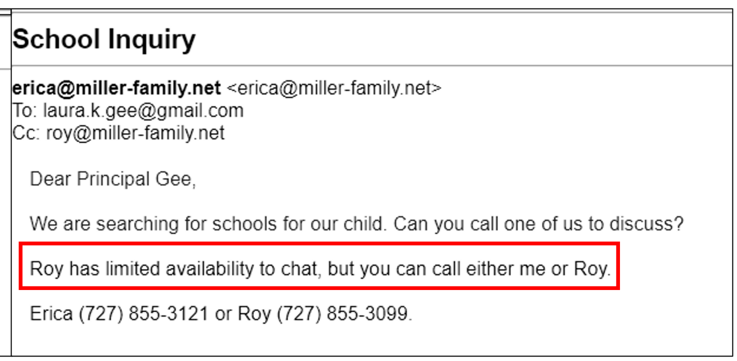
roy@miller-family.net <roy@miller-family.net>
To: laura.k.gee@gmail.com
Cc: erica@miller-family.net

Dear Principal Gee,

We are searching for schools for our child. Can you call one of us to discuss?

I have a lot of availability to chat, but you can call either me or Erica.

Roy (727) 855-3143 or Erica (727) 855-3100.



School Inquiry

erica@miller-family.net <erica@miller-family.net>
To: laura.k.gee@gmail.com
Cc: roy@miller-family.net

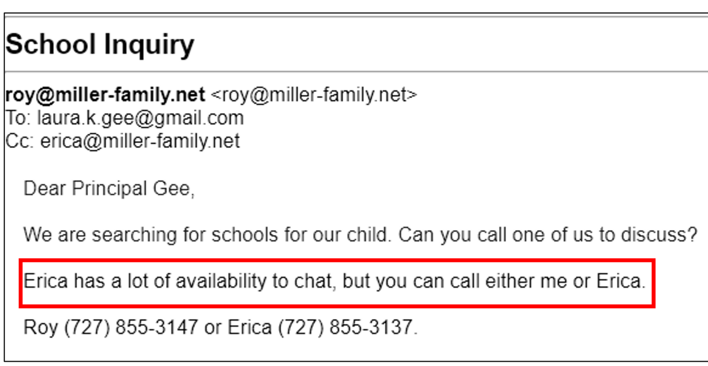
Dear Principal Gee,

We are searching for schools for our child. Can you call one of us to discuss?

Roy has limited availability to chat, but you can call either me or Roy.

Erica (727) 855-3121 or Roy (727) 855-3099.

Figure F.3: Main: High Female and Low Female Signal



School Inquiry

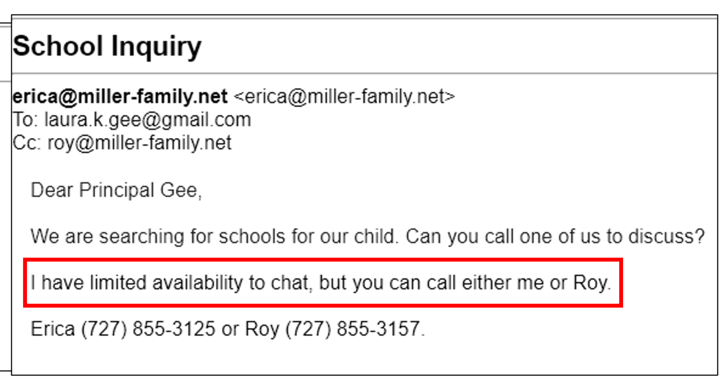
roy@miller-family.net <roy@miller-family.net>
To: laura.k.gee@gmail.com
Cc: erica@miller-family.net

Dear Principal Gee,

We are searching for schools for our child. Can you call one of us to discuss?

Erica has a lot of availability to chat, but you can call either me or Erica.

Roy (727) 855-3147 or Erica (727) 855-3137.



School Inquiry

erica@miller-family.net <erica@miller-family.net>
To: laura.k.gee@gmail.com
Cc: roy@miller-family.net

Dear Principal Gee,

We are searching for schools for our child. Can you call one of us to discuss?

I have limited availability to chat, but you can call either me or Roy.

Erica (727) 855-3125 or Roy (727) 855-3157.

G Theory Appendix

G.1 Notation

We provide a summary of our notation as a reference.

Sub- and superscripts

- $i \in I$: decision maker subscript
- $j \in \{n, f, m\}$: subscript for which parent to call first
- $t \in \{\boldsymbol{b}aseline, \boldsymbol{h}igh\boldsymbol{F}emale, \boldsymbol{l}ow\boldsymbol{F}emale, \boldsymbol{h}igh\boldsymbol{M}ale, \boldsymbol{l}ow\boldsymbol{M}ale\}$: treatment superscript. When it is only relevant that a message was sent about a particular parent (not whether it was low or high), we use $\boldsymbol{M}$ and $\boldsymbol{F}$
- $g \in \{R, N\}$: second superscript for principal characteristic

Objects of interest

1. Structural parameters: $\delta, r, \omega^2, \lambda$
   - e.g. δ_m^R for the deterrents principals of religious schools face to calling male parent
2. Reduced form parameters: α, η, γ
   - e.g. $\gamma_m^{\boldsymbol{F},\, R}$ for impact of signal about female parent on probability that principal from religious school will call male parents
3. Reduced-form regressors: w and x do not vary with principal characteristics, so we have $w_{i,m}^{h\boldsymbol{F}} = 0$ and $x_{i,m}^{h\boldsymbol{F}} = 0$ for the impact on principal valuation of calling the male parent when they receive a high signal about the female parent
4. Proportions of decision makers: $p_m^{\boldsymbol{F},\, R}$
5. Coefficients in treatment effects regression: $\beta^{l\boldsymbol{M}}, \beta^{h\boldsymbol{M}}, \beta^{l\boldsymbol{F}}, \beta^{h\boldsymbol{F}}$
   - e.g. $\beta_m^{l\boldsymbol{F},\, R}$ for impact of low signal about female parent on the probability that a religious-school principal will call the male parent

G.2 Base Theoretical Framework

G.2.1 Proof of Result 1 (Identification of Reduced Form Parameters)

In Sections 2.1–2.3, we assume the following:

1. Decision maker i chooses from among three alternatives: $j \in \{n, f, m\}$.
2. Decision maker i holds probabilistic beliefs about the probability that alternative j will respond to a phone call, $r_{ij} \sim \mathcal{N}(\bar{r}_j, \omega_j^2)$.
3. $r_{i,f}$ and $r_{i,m}$ are independent.
4. Decision makers are risk neutral.[1]
5. Each decision maker faces a cost c_i for making a call that is the same for alternatives f and m. c is the population mean of c_i.
6. Each decision maker has a non-belief deterrent parameter for calling that varies by alternative.
7. Each decision maker i knows c_i and δ_{ij}.
8. Expected utility for decision maker i is $\mathbb{E}(U_{ij}) = \mathbb{E}(r_{ij}) - \big(\delta_{ij} + c_i\big)$ for $j \in \{n, f, m\}$ with $\mathbb{E}(U_{i,n}) = 0$.
9. The experimenters choose signal values x_{ij} at random to show each decision maker and send a signal $x_{ij} \in \{-1, 1\}$ about at most one alternative to each decision maker. The decision makers believe that $x_{ij} \sim \mathcal{N}(r_j, \sigma^2)$, $j \in \{f, m\}$, where r_j is the true responsiveness of j.
10. A signal x_{ij} can shift the belief $\tilde{r}_{ij}$ but does not affect c_i or δ_{ij}.
11. ε_{ij} are each distributed according to the standard Gumbel distribution.

Given the above assumptions and the experimental data, we can use the observable proportions of decision makers in each signal-outcome pair to identify the reduced-form parameters.

We begin with the case in which no signal is sent about either alternative, i.e. $w_{ij} = 0 \; \forall j$. Here, the terms involving η_j and γ_j are zero for all decision makers, so we have $U_{ij} = \alpha_j \; \forall j$. Because $U_{i,n} = \alpha_n = 0$ by assumption, the probabilities from the logit model are

$$p_n^b \equiv \frac{1}{1 + e^{\alpha_f} + e^{\alpha_m}} \quad p_f^b \equiv \frac{e^{\alpha_f}}{1 + e^{\alpha_f} + e^{\alpha_m}} \quad p_m^b \equiv \frac{e^{\alpha_m}}{1 + e^{\alpha_f} + e^{\alpha_m}}$$

[1] We have assumed that decision makers are risk neutral with respect to the decision about whether and whom to call. In Appendix G.5, we discuss relaxing this assumption.

where subscripts denote which alternative is chosen and the superscript b denotes that no signal is sent about either alternative (this is our baseline treatment).

Sending a signal ($w_{i,f} = 1$) with value $x_{i,f} = 1$ about alternative f and no signal about alternative m makes the deterministic part of utility for alternative f (i.e. Equation 4 without the error) $\alpha_f + \eta_f + \gamma_f$. We therefore have the following probabilities:

$$p_n^{h\boldsymbol{F}} \equiv \frac{1}{1 + e^{\alpha_f + \eta_f + \gamma_f} + e^{\alpha_m}} \quad p_f^{h\boldsymbol{F}} \equiv \frac{e^{\alpha_f + \eta_f + \gamma_f}}{1 + e^{\alpha_f + \eta_f + \gamma_f} + e^{\alpha_m}} \quad p_m^{h\boldsymbol{F}} \equiv \frac{e^{\alpha_m}}{1 + e^{\alpha_f + \eta_f + \gamma_f} + e^{\alpha_m}}$$

where the superscript "$h\boldsymbol{F}$" denotes that we send only a high signal (i.e. value of 1) about alternative f.

Similarly, when we send a signal with value $x_{i,f} = -1$ about alternative f and no signal about alternative m makes the deterministic part of utility for alternative f (i.e. Equation 4 without the error) $\alpha_f + \eta_f - \gamma_f$. We therefore have the following probabilities:

$$p_n^{l\boldsymbol{F}} \equiv \frac{1}{1 + e^{\alpha_f + \eta_f - \gamma_f} + e^{\alpha_m}} \quad p_f^{l\boldsymbol{F}} \equiv \frac{e^{\alpha_f + \eta_f - \gamma_f}}{1 + e^{\alpha_f + \eta_f - \gamma_f} + e^{\alpha_m}} \quad p_m^{l\boldsymbol{F}} \equiv \frac{e^{\alpha_m}}{1 + e^{\alpha_f + \eta_f - \gamma_f} + e^{\alpha_m}}$$

where the superscript "$l\boldsymbol{F}$" denotes that we send only a low signal (i.e. value of -1) about alternative f.

We repeat each of the last two conditions for alternative m. Sending a signal ($w_{i,m} = 1$) with value $x_{i,m} = 1$ about alternative m and no signal about alternative f leads to the following probabilities:

$$p_n^{h\boldsymbol{M}} \equiv \frac{1}{1 + e^{\alpha_f} + e^{\alpha_m + \eta_m + \gamma_m}} \quad p_f^{h\boldsymbol{M}} \equiv \frac{e^{\alpha_f}}{1 + e^{\alpha_f} + e^{\alpha_m + \eta_m + \gamma_m}} \quad p_m^{h\boldsymbol{M}} \equiv \frac{e^{\alpha_m + \eta_m + \gamma_m}}{1 + e^{\alpha_f} + e^{\alpha_m + \eta_m + \gamma_m}}$$

Sending a signal with value $x_{i,m} = -1$ about alternative m and no signal about alternative f leads to the following probabilities:

$$p_n^{l\boldsymbol{M}} \equiv \frac{1}{1 + e^{\alpha_f} + e^{\alpha_m + \eta_m - \gamma_m}} \quad p_f^{l\boldsymbol{M}} \equiv \frac{e^{\alpha_f}}{1 + e^{\alpha_f} + e^{\alpha_m + \eta_m - \gamma_m}} \quad p_m^{l\boldsymbol{M}} \equiv \frac{e^{\alpha_m + \eta_m - \gamma_m}}{1 + e^{\alpha_f} + e^{\alpha_m + \eta_m - \gamma_m}}$$

Next, we manipulate the logit probabilities to identify reduced-form parameters $\alpha_j, \eta_j, \gamma_j$. In order to identify α_j, we take ratios of the probabilities for when no signal is sent.

$$\frac{p_j^b}{p_n^b} = e^{\alpha_j} \Leftrightarrow \boxed{\alpha_j = \ln p_j^b - \ln p_n^b \text{ for } j \in \{f, m\}} \tag{11}$$

In order to identify η_j and γ_j, we must combine equations. Start with

$$\frac{p_j^{hJ}}{p_n^{hJ}} = e^{\alpha_j + \eta_j + \gamma_j} \Leftrightarrow \alpha_j + \eta_j + \gamma_j = \ln p_j^{hJ} - \ln p_n^{hJ} \tag{12}$$

and

$$\frac{p_j^{lJ}}{p_n^{lJ}} = e^{\alpha_j+\eta_j-\gamma_j} \Leftrightarrow \alpha_j + \eta_j - \gamma_j = \ln p_j^{lJ} - \ln p_n^{lJ} \tag{13}$$

where $J \in \{F, M\}$ denotes about which parent the signal is sent.

Subtracting Equation (13) from Equation (12), we have

$$\alpha_j + \eta_j + \gamma_j - \alpha_j - \eta_j + \gamma_j = \ln p_j^{hJ} - \ln p_n^{hJ} - \ln p_j^{lJ} + \ln p_n^{lJ}$$

Simplifying, we have

$$2\gamma_j = \ln p_j^{hJ} - \ln p_n^{hJ} - \ln p_j^{lJ} + \ln p_n^{lJ} \Leftrightarrow \boxed{\gamma_j = \tfrac{1}{2}\left[\ln p_j^{hJ} - \ln p_n^{hJ} - \ln p_j^{lJ} + \ln p_n^{lJ}\right] \text{ for } j \in \{f, m\}} \tag{14}$$

Combining Equations (11), (12) and (13), we have

$$\ln p_j^b - \ln p_n^b + \eta_j + \frac{1}{2}\left[\ln p_j^{hJ} - \ln p_n^{hJ} - \ln p_j^{lJ} + \ln p_n^{lJ}\right] = \ln p_j^{hJ} - \ln p_n^{hJ}$$

Simplifying

$$\eta_j = -\ln p_j^b + \ln p_n^b + \ln p_j^{hJ} - \ln p_n^{hJ} - \frac{1}{2}\left[\ln p_j^{hJ} - \ln p_n^{hJ} - \ln p_j^{lJ} + \ln p_n^{lJ}\right]$$

$$\boxed{\eta_j = -\ln p_j^b + \ln p_n^b + \tfrac{1}{2}\left[\ln p_j^{hJ} - \ln p_n^{hJ} + \ln p_j^{lJ} - \ln p_n^{lJ}\right] \text{ for } j \in \{f, m\}} \tag{15}$$

■

It is worth noting that η_j and γ_j cannot vary independently: every term in Expression (14) is also present in Expression (15).

G.2.2 Proof of Result 2 (Identification of Structural Parameters)

We use the six identified reduced-form parameters from Result 1 to identify the deep parameters $\lambda_f, \lambda_m, \bar{r}_f, \bar{r}_m$ and $\bar{\delta}_f - \bar{\delta}_m$. Recall Equations (5)-(7) that relate the reduced-form parameters to the deep parameters:

$$\alpha_j = \bar{r}_j - \bar{\delta}_j - c$$
$$\eta_j = -(1 - \lambda_j)\bar{r}_j$$
$$\gamma_j = 1 - \lambda_j$$

We can directly identify $\lambda_j = 1 - \gamma_j$ using the third of these equations. Recall that λ_j is composed of σ^2 and ω_j^2, but these can't be separately identified since we do not vary ω_j^2.

We can then combine the equation for λ_j with the second equation to find $\bar{r}_j = -\frac{\eta_j}{\gamma_j}$. Plugging this into the first equation produces $\bar{\delta}_j + c = -\frac{\eta_j}{\gamma_j} - \alpha_j$.

We cannot separately identify $\bar{\delta}_f$ and $\bar{\delta}_m$, but we can combine $\bar{\delta}_f + c = -\frac{\eta_f}{\gamma_f} - \alpha_f$ and $\bar{\delta}_m + c = -\frac{\eta_m}{\gamma_m} - \alpha_m$ from the previous step to get $\bar{\delta}_m - \bar{\delta}_f = \frac{\eta_f}{\gamma_f} - \frac{\eta_m}{\gamma_m} + \alpha_f - \alpha_m$. ■

G.3 Model with decision maker characteristics

We let g index the discrete categories that make up the decision maker characteristic. Each type g of the decision maker makes their decision as in Section 2.3. This model extension applies to any observable characteristic of decision makers that is discrete in nature. Here, we focus on the type of school at which the decision maker works so that $G = \{R, N\}$, where decision makers at religious schools are denoted by R and decision makers at non-religious schools are denoted by N.

With decision maker characteristics, Equation 1 becomes

$$\mathbb{E}(U_{ij}^g) = \mathbb{E}(r_{ij}^g) - \delta_{ij}^g - c_i$$

Each type g of the decision maker makes their decision as in Section 2.3. The signals about parental responsiveness are not differentiated by type of principal, but the signals may have differential impact on the beliefs of different types. We extend the assumptions of Section 2.3 so that beliefs are not only independent across alternatives but also across types of decision maker, i.e. that all $r_{ij}^g \sim \mathcal{N}(\bar{r}_j^g, \omega_j^2)$ are mutually independent.[2]

We now have that decision maker i of type g has a posterior mean $\tilde{r}_{ij}^g$ for the responsiveness of j, assuming the prior variance is *common* to all i, is

$$\tilde{r}_{ij}^g = \lambda_j^g \bar{r}_{ij}^g + (1 - \lambda_j^g) x_{ij}, \quad \lambda_j^g = \frac{1/(\omega_j^g)^2}{1/(\omega_j^g)^2 + 1/\sigma^2}.$$

We continue to assume that decision-maker beliefs are not heterogeneous within type, so that $\bar{r}_{ij}^g = \bar{r}_j^g \ \forall i$. Since signals are not differentiated by decision maker type, Equation 3 becomes

$$\mathbb{E}(U_{ij}^g) = \bar{r}_j^g - (1 - \lambda_j^g)\bar{r}_j^g w_{ij} + (1 - \lambda_j^g) w_{ij} x_{ij} - \delta_{ij}^g - c_i$$

for all j and g.

[2]This assumption of independence will be relaxed in the next section.

Equations (4)-(8) become

$$U^g_{ij} = \alpha^g_j + \eta^g_j w_{ij} + \gamma^g_j w_{ij} x_{ij} + \varepsilon^g_{ij} \quad (16)$$
$$\alpha^g_j = \bar{r}^g_j - \bar{\delta}^g_j - c \quad (17)$$
$$\eta^g_j = -(1 - \lambda^g_j)\bar{r}^g_j \quad (18)$$
$$\gamma^g_j = 1 - \lambda^g_j \quad (19)$$
$$\varepsilon^g_{ij} = (c - c_i) + (\bar{\delta}^g_j - \delta^g_{ij}) \quad (20)$$

where $\bar{\delta}^g_j$ denotes the average value of δ^g_{ij}.

We then have the following identification result:

Result 4. *Given the assumptions of Sections 2.1–2.4 and this section, the reduced-form parameters* α^g_j, γ^g_j, η^g_j *and the structural parameters* $\lambda^g_f, \lambda^g_m, \bar{r}^g_f, \bar{r}^g_m$, $\bar{\delta}^g_f - \bar{\delta}^g_m$ *are identified for* $j \in \{f, m\}$ *and* $g \in G$*, G discrete.*

Proof: Repeatedly apply the proofs for Results 1 and 2 for each $g \in G$. ■

G.3.1 Testable Hypotheses

Result 4 allows us to identify beliefs about parents' responsiveness versus other deterrents for each type of decision maker. The testable hypotheses from Section 2.5 can again be tested here, with one version for each characteristic of the decision makers. Here we focus on the impact of gender norms, hypothesizing that religious schools may have more traditional gender norms.

Hypothesis 4. *At baseline (i.e., when no signal is sent about availability), decision makers at religious schools are more likely to call a female parent and less likely to call a male parent than decision makers at non-religious schools. That is, the proportion of calls to female parents will be higher when the decision maker is from a religious school, and the proportion of calls to male parents will be higher when the decision maker is from a non-religious school. We find support for this hypothesis if* $p^{b,R}_f > p^{b,N}_f$ *and* $p^{b,N}_m > p^{b,R}_m$ *where decision maker's type is represented by superscripts R and N.*

This is equivalent to $\alpha^R_f > \alpha^N_f$ and $\alpha^N_m > \alpha^R_m$ in terms of the reduced-form parameters.

Hypothesis 5. *The relative deterrents to calling male parents versus female parents is larger for decision makers at religious schools than it is for decision makers at non-religious schools. We find support for this hypothesis if* $\bar{\delta}^R_m - \bar{\delta}^R_f > \bar{\delta}^N_m - \bar{\delta}^N_f$.

G.4 Relaxing the independence assumption

We now assume (1) the distributions of the signals about the two parents can not only have different means but also different variances and (2) the impact on decision maker updating can be summarized by a correlation parameter ρ_j, which captures the impact on the belief about parent j from a signal about the other parent.

In addition to the new structural parameters ρ_j, this version of the model also has update parameters λ_j^t that are differentiated not only by the parent about whom the update is being made (j), but now also by the parent about whom the message is sent (t). The reduced form parameters η and γ are also now differentiated by the parent about whom the signal is sent.

After relaxing the assumption that a signal about one parent only affects the belief about that parent, the updating process becomes more complex. Note that, in order to keep notation simple, we focus without loss of generality on how the belief about the female parent is updated. There now need to be two versions of Equation (2):

$$\widetilde{r}_{if}^F = \lambda_f^F \bar{r}_f + (1-\lambda_f^F)x_{if}, \quad \lambda_f^F = \frac{1/\omega_f^2}{1/\omega_f^2 + 1/\sigma_F^2} \tag{21}$$

$$\widetilde{r}_{if}^M = \lambda_f^M \bar{r}_f + (1-\lambda_f^M)\rho_f x_{im}, \quad \lambda_f^M = \frac{1/\omega_f^2}{1/\omega_f^2 + 1/\sigma_M^2} \tag{22}$$

with

$$\bar{r}_f \sim \mathcal{N}(r_f, \omega_f^2), \quad x_{if} \sim \mathcal{N}(r_f, \sigma_F^2), \quad x_{im} \sim \mathcal{N}(r_m, \sigma_M^2).$$

We now have two ways that decision maker i's belief about the female parent can be updated: via a signal directly about the female parent ($w_{i,f} = 1$ and $w_{i,m} = 0$), or via a signal about the male parent ($w_{i,f} = 0$ and $w_{i,m} = 1$).[3]

Under this more general formulation, Equation (3) becomes

$$\begin{aligned}
\mathbb{E}(U_{i,f}) &= (1 - w_{i,f} - w_{i,m})\bar{r}_f + w_{i,f}\widetilde{r}_{i,f}^F(x_{i,f}, x_{i,m}) + w_{i,m}\widetilde{r}_{i,f}^M(x_{i,f}, x_{i,m}) - (\delta_{ij} + c_i) && (23)\\
&= (1 - w_{i,f} - w_{i,m})\bar{r}_f + w_{i,f}\left[\lambda_f^F \bar{r}_f + (1-\lambda_f^F)x_{i,f}\right] + && (24)\\
&\qquad w_{i,m}\left[\lambda_f^M \bar{r}_f + (1-\lambda_f^M)(\rho_f)x_{i,m}\right] - (\delta_{ij} + c_i) && (25)\\
&= \bar{r}_j - (1-\lambda_f^F)\bar{r}_f w_{i,f} - (1-\lambda_f^M)\bar{r}_f w_{i,m} + && (26)\\
&\qquad (1-\lambda_f^F)w_{i,f}x_{i,f} + (1-\lambda_f^M)\rho_f w_{i,m}x_{i,m} - (\delta_{ij} + c_i) && (27)\\
&= \alpha_f + \eta_f^F w_{i,f} + \eta_f^M w_{i,m} + \gamma_f^F w_{i,f}x_{i,f} + \gamma_f^M w_{i,m}x_{i,m} + \varepsilon_{i,f}. && (28)
\end{aligned}$$

[3]This formulation is relatively simple because we only send signals about one parent to any given decision maker. It can be generalized for the case where one sends signals about both parents to the same decision maker.

where the last equation follows from the mapping below:

$$\begin{aligned}
\alpha_f &= \bar{r}_f - \bar{\delta}_f - c \\
\eta_f^F &= -(1-\lambda_f^F)\bar{r}_f \\
\eta_f^M &= -(1-\lambda_f^M)\bar{r}_f \\
\gamma_f^F &= 1-\lambda_f^F \\
\gamma_f^M &= (1-\lambda_f^M)\rho_f \\
\varepsilon_{i,f} &= (c-c_i)+(\bar{\delta}_f-\delta_{i,f}).
\end{aligned}$$

Result 5. *Given the assumptions of Sections 2.1–2.6.3, the reduced-form parameters α_j, γ_j^t, η_j^t and the structural parameters $\lambda_j^t, \rho_j, \bar{r}_j$ and $\bar{\delta}_f - \bar{\delta}_m$ are identified for $j \in \{f, m\}$ and $t \in \{F, M\}$.*

Proof: α_f is once again identified by Equation (5). Equations (6) and (7) identify η_f^F and γ_f^F (with only a notational change from η_f and γ_f to η_f^F and γ_f^F). The following equations identify η_f^M and γ_f^M.

$$\frac{p_f^{hM}}{p_n^{hM}} = e^{\alpha_f+\eta_f^M+\gamma_f^M\rho_f} \Leftrightarrow \alpha_f+\eta_f^M+\gamma_f^M\rho_f = \ln\left(\frac{p_f^{hM}}{p_n^{hM}}\right) \Leftrightarrow \alpha_f+\eta_f^M+\gamma_f^M\rho_f = \ln p_f^{hM} - \ln p_n^{hM} \tag{29}$$

$$\frac{p_f^{lM}}{p_n^{lM}} = e^{\alpha_f+\eta_f^M-\gamma_f^M\rho_f} \Leftrightarrow \alpha_f+\eta_f^M-\gamma_f^M\rho_f = \ln\left(\frac{p_f^{lM}}{p_n^{lM}}\right) \Leftrightarrow \alpha_f+\eta_f^M-\gamma_f^M\rho_f = \ln p_f^{lM} - \ln p_n^{lM} \tag{30}$$

Subtracting Equation (30) from Equation (29), we have

$$\alpha_f+\eta_f^M+\gamma_f^M\rho_f-\alpha_f-\eta_f^M+\gamma_f^M\rho_f = \ln p_f^{hM}-\ln p_n^{hM}-\ln p_f^{lM}+\ln p_n^{lM}$$

Simplifying, we have

$$2\rho_f\gamma_f^M = \ln p_f^{hM}-\ln p_n^{hM}-\ln p_f^{lM}+\ln p_n^{lM} \Leftrightarrow \boxed{\rho_f\gamma_f^M = \tfrac{1}{2}\left[\ln p_f^{hM}-\ln p_n^{hM}-\ln p_f^{lM}+\ln p_n^{lM}\right]} \tag{31}$$

Combining Equations (11), (29) and (30), we have

$$\ln p_f^b - \ln p_n^b + \eta_f^M + \frac{1}{2}\left[\ln p_f^{hM}-\ln p_n^{hM}-\ln p_f^{lM}+\ln p_n^{lM}\right] = \ln p_f^{hm} - \ln p_n^{hM}$$

Simplifying

$$\eta_f^M = -\ln p_f^b + \ln p_n^b + \ln p_f^{hM} - \ln p_n^{hM} - \frac{1}{2}\left[\ln p_f^{hM}-\ln p_n^{hM}-\ln p_f^{lM}+\ln p_n^{lM}\right]$$

$$\boxed{\eta_f^M = -\ln p_f^b + \ln p_n^b + \tfrac{1}{2}\left[\ln p_f^{hM} - \ln p_n^{hM} + \ln p_f^{lM} - \ln p_n^{lM}\right]} \tag{32}$$

Analogous equations similarly identify η_m^F and γ_m^F. Combined with the results for the male parent in the proof of Result 1, all reduced-form parameters in the model generalized for correlations are identified.

It is left to show that the structural parameters are identified. Again, it is without loss of generality to demonstrate identification for the parameters about the female parent; analogous equations for the male parent hold.

As in the proof of Result 2, γ_f^F directly identifies λ_f^F as $\lambda_f^F = 1 - \gamma_f^F$. Once we have λ_f^F, we combine it with the η_f^F equation to get $\bar{r}_f = -\frac{\eta_f^F}{\gamma_f^F}$. Next, we use the η_f^M equation to get $\lambda_f^M = 1 - \frac{\eta_f^M}{\eta_f^F}\gamma_f^F$. We now have everything we need to derive $\rho_f = \frac{\eta_f^F}{\gamma_f^F} \cdot \frac{\gamma_f^M}{\eta_f^M}$ from the γ_f^M equation. Finally, from the α_f equation, we have $\bar{\delta}_f + c = -\frac{\eta_f^F}{\gamma_f^F} - \alpha_f$. Subtracting this equation from $\bar{\delta}_m + c = -\frac{\eta_m^M}{\gamma_m^M} - \alpha_m$, we have $\bar{\delta}_m - \bar{\delta}_f = \frac{\eta_f^F}{\gamma_f^F} - \frac{\eta_m^M}{\gamma_m^M} + \alpha_f - \alpha_m$ as in Result 2. ■

Careful examination of the proof of Result 5 compared to those of Results 1 and 2 will make clear that the identification of the parameters in the base model is not disturbed by a correlation in the belief updating process. This is because identification of those parameters only involves the number of calls to parent j versus neither parent after a signal about parent j compared to the baseline message. Generalizing the model to allow for this correlation simply lets us test the independence assumption and then to quantify the size of the correlation and any potential differences in the updating processes after signals about male versus female parents.

G.4.1 Testable Hypothesis

Now that we have established that both the reduced-form and structural parameters are well-identified, we put forward an additional testable hypothesis that emerges from the extended model. Note that the testable hypotheses from Section 2.5 remain valid under this model extension.

Hypothesis 6. *Decision makers infer information:*

i) about the male parent after receiving a signal about the female parent; and

ii) about the female parent after receiving a signal about the male parent.

We find support for this hypothesis if (i) $\rho_m \neq 0$ and (ii) $\rho_f \neq 0$.

The interpretation of the sign of ρ_j is as follows: If ρ_j is positive, a positive (negative) signal about parent j is taken to also be a positive (negative signal about the other parent. If ρ_j is negative, a positive (negative) signal about parent j is taken to be a *negative (positive)* signal about the other parent. Hypothesis 6 simply says that decision makers infer information about both parents from a signal about one parent.

G.5 Risk Aversion

We have assumed that decision makers are risk neutral with respect to the decision about whether and whom to call. If decision makers are instead risk averse with respect to this decision, the prior variance will play a role in the outcome. Importantly, risk-averse decision makers who are less uncertain about female parents have an additional reason to call female parents beyond their average beliefs.

In terms of the identification of our parameters, what we attribute entirely to the mean of the belief distribution is actually a combination of the mean and the variance if decision makers are risk averse. In this case, the parameter we estimate for the mean belief about female parents could be larger than the actual mean belief. If, instead, decision makers are more uncertain about female parents, our estimated belief about the female parent will be smaller than the actual mean belief. The implications for the belief about the male parent mirror these relationships.

To develop intuition for the effect of risk aversion, imagine that a decision maker holds the same beliefs and has the same other deterrents parameter for both parents. This decision maker will call the parent about whom she is less uncertain; that is, she calls the parent for whom her updated belief variance is smaller. Given a signal variance that is common to both parents, the updated belief variance is lower for the parent about whom the prior belief variance is lower.

We can infer the ordering of the prior belief variance by comparing the weights that decision makers place on the prior belief, λ_m and λ_f. Assume without loss of generality that decision makers place greater weight on the prior variance for the female parent, that is $\lambda_f > \lambda_m$. This implies that the prior and posterior variance for the belief about females is lower, i.e. $\omega_f^2 < \omega_m^2$. Intuitively, decision makers place less weight on the prior belief when the prior belief is more uncertain.

H External Validity

H.1 Type of Household

The primary goal of our work is to identify gender gaps in households with two parents where one identifies as female and the other as male. We fully acknowledge that gender identity takes more than two values, but we have started this research with the two ends of the gender spectrum (male and female).

About 98% of US persons identify as either male or female, with the remaining 2% identifying in a number of different ways.[4] The plurality of households with children under the age of 18, 84%, live in a home with two parents – with 99% of these being opposite gender couples.[5]

We believe the direction of the effects of our high/low-availability messages would be the same for a variety of genders (e.g. two non-binary parents, same-sex couples), however we would expect baseline inequality to be closer to zero in households with these gender identity sets. And, indeed nationally representative data indicates that same-sex households do not report wishing they were contacted more or less than they actually are by their child's school.[6]

H.2 School Setting

Our experiment takes place in a K–12 school setting which we chose because over 40% of households in the US, have school-aged children (NCES, 2021). Almost all, 97%, of parents send their children to school outside the home (NCES, 2021). Additionally, the gender gap in time spent on children in school-related activities closely mirrors the overall tendency for mothers to engage in more child-related tasks than fathers (BLS, 2021).

We believe that any gender gaps that we document in our specific task in the school setting will generalize to other tasks in the school setting, such as picking up a sick child, or joining the Parent Teacher Association. Educators in our survey report that they would favor contacting the mother first in many of these scenarios (we discuss the survey in Section K). The gender distribution of these tasks is significantly skewed with mothers comprising almost 90% of Parent Teacher Association members and many surveys finding fathers self-report lower levels of involvement in their child's school activities, compared to mothers.[7]

Furthermore, although the gender inequality that we document is in the school setting, this is only one of many settings where mothers spend significantly more time on children than fathers. Prior studies have documented substantial gender differences in time devoted to caring for sick children, taking children to the doctor, and coordinating a wide range of household and child-related tasks, known as cognitive labor.[8] If these other inequalities are

4 https://www.census.gov/library/stories/2021/11/census-bureau-survey-explores-sexual-orientation-and-g html

5 https://www.census.gov/library/stories/2022/07/most-kids-with-parent-in-same-sex-relationship-live-w html

6 See https://csed.byu.edu/american-family-survey for evidence from 205 respondents who are nationally representative. The limited survey evidence we have on non-binary parents from this survey does indicate that the three non-binary respondents report being contacted 75% but wishing to be contacted only 67% of the time. See also our own survey in section K.2 and

7 See our own survey in section K.2 and Daly and Groes (2017) https://archive.nytimes.com/parenting.blogs.nytimes.com/2009/01/06/dads-in-the-pta/, https://education.gov.scot/media/b3cn2mv5/nih327-dads-involvement-in-school.pdf

8 Wikle and Cullen (2023); Bianchi et al. (2006); Boye (2015); Daly and Groes (2017); Daminger (2019); **?**); Charmes (2019) https://www.bls.gov/spotlight/2017/differences-in-parents-time-use-between-the-summer-and-the-school-year/home.htm https://melbourneinstitute.unimelb.edu.au/__data/assets/pdf_file/0009/3963249/HILDA-Statistical-Report-2021.pdf

also partially driven by external demands, our findings likely represent a lower bound for the overall gender gaps in external demands for parental involvement.

I Ethics

A common critique of audit studies, which perform outreach from fictitious persons to a third party (often a business that is hiring), is that the person who receives the message wastes time and effort on evaluating the message. We estimate that the median time spent leaving our parents a message was 50 seconds, with the 99th percentile being a message of less than two minutes. As such, each principal in our dataset is not spending a large amount of time being in our study. Furthermore, unlike a resume audit study, the principals in our study do not need to evaluate a lengthy fictitious candidate's resume for a position, rather they need only to read our brief email message and return our call (only 20% of principals call us, and only 17% leave a voice mail, further reducing the likelihood of significant harm to our subjects).

We considered writing positive reviews for schools as a form of compensation for their time, but after consultation with our IRB, were told this would likely be a violation of the terms of service of the review websites, and as such, we could not gain IRB approval for this.

Also, our subjects are school officials who as part of their position aim to provide increased school quality. Our research, in part, informs ways to increase school quality through better serving parents, and as such, participation in our study is part of our subjects' regular job duties.

A second concern is that the decision makers' involvement may harm other non-fictitious persons because of their involvement in the audit study. For example, if a firm decides to call back a fictitious applicant in an audit study, this may crowd out a call to a real applicant. We do not believe our study poses this harm. The act of calling one family likely does not crowd out further actions.

An additional possible hazzard in a labor-market audit study: the fictitious applicants never accept the job interviews, and if they have some identifiable factor, such as foreign sounding names, this may cause firms to negatively update their views of real persons with foreign sounding names. Again, we do not think our study poses this harm as all of our households are two-parent households with racially neutral names, as such it is difficult to identify which subgroup a school principal would negatively update about in response to our study.

Lastly, a large survey of economists finds that researchers are quite comfortable with the lack of informed consent common in natural field experiments like audit studies (Charness et al., 2022). The same survey finds that economists prioritize avoiding more explicit deception but believe it is acceptable for important questions when alternative research designs are unavailable. Informed consent is ideal, but it is difficult to study gender discrimination with informed consent without possibly biasing the results. Our study was approved by the relevant Institutional Review Boards (IRBs) at our home institutions, and as such the harms and benefits have been evaluated and approved by a third party.

J Data Collection and Matching

J.1 Emails and Phone Numbers

To record phone metadata and voicemails we used a service called Callfire to set up a series of different phone numbers for our male and female parents. First, we set up a series of phone numbers with a generic voice mail box and text-message auto-reply saying that number did not receive text messages. We also set up email addresses with an auto-reply directing responders to please call instead of emailing. The exact email addresses that we sent our messages from were "erica@miller-family.net" and "roy@miller-family.net" for part of our data collection. We switched to emails from "audrey@the-johnsonfamily.net" and "curtis@the-johnsonfamily.net" for the bulk of data collection. We discuss the choice of exact names in detail below and in Section J.4. Due to constraints on email send limits, the follow-up emails sent about two weeks after the first email which said the family no longer needed to talk were sent from "audrey@the-johnson-family.net " and "curtis@the-johnson-family.net."

Email is a common way for parents to contact schools. In our own survey, three-fourths of educators reported being contacted by parents via email at least once a month (Section K). These educators also reported that, when being emailed by both parents, a single parent emailing and cc'ing the other parent was more common than emails from a joint family email account. In one of our pilot data-collection efforts, we found that emailing from a joint email account lowered callback rates (Section J.4). Furthermore, we were concerned that a joint family email address might signal a more egalitarian family, which might bias our results towards finding more equal calls to mothers and fathers. As such, we decided to not use any joint family email accounts. We note, however, that this choice differs from our original pre-registration, in which we proposed sending emails from a joint family email account.[9]

J.2 Names

We chose the names from the top 200 listed by the Social Security Administration in 1980.[10] We chose 1980 because we primarily contact schools that enroll children ages 5 to 18, the average age being 11.5 years old. A child who is 11.5 now was born in 2009 (2021-11.5=2009.5). The average age of a first-time parent in 2009 was 29.4 years old,[11] which means our parents on average would have been born in 1980 (because 2009-29.4=1979.6). From the 1980 list, we chose first names that did not have a strong indication of a specific race or ethnicity

[9] https://www.socialscienceregistry.org/trials/7610

[10] https://www.ssa.gov/OACT/babynames/decades/names1980s.html

[11] CBS News "Average age of first-time mothers up to 29.9 years," November 5, 2019 https://www.cbs.nl/en-gb/news/2019/19/average-age-of-first-time-mothers-up-to-29-9-years

Table J.1: Longer Versions of Messages

Variation & Treatment	Body Text
Main Baseline (Used in Study)	We are searching for schools for our child. Can you call one of us to discuss?
Main Baseline (Longer Alternative)	I'm Curtis[Audrey] Johnson. I'm writing to request information about your school because we are searching for schools for our child, Riley. Riley is a well behaved student, and loves most subjects. We're not totally sure when we will be needing to enroll, but we are looking forward to hearing more from you at your earliest convenience. Could you call one of us to discuss? Thank you very much,
Equal Decision (Used in Study)	We are searching for schools for our child. Can you call one of us to discuss? This is the type of decision we both want to be involved in equally.
Equal Decision (Longer Alternative)	We are searching for schools for our child. Could you call one of us to discuss? You can call either me or my wife, Audrey [husband, Curtis]. Since we make these kinds of decisions together, whoever you call will convey the information to the other parent. Thank you very much,

(Tzioumis, 2018) (Erica and Roy) and we chose our last names (Johnson and Miller) from the list of the most common last names in the US over many decades.[12]. We also did online searches for the names (Audrey Johnson, Curtis Johnson, Erica Miller, Roy Miller) to see if there were any famous or infamous people with these names that might bias our results.[13] In addition we did a Google image search for these names to ensure they encompassed a balance of race and ethnicities.

J.3 Messages

We pretested our messages using a survey run on Amazon Mechanical Turk to select which messages gave the widest variation in self-reported likelihood of getting a call back. We also pretested our messages with a set of educators (see Section K) to ensure the messages seemed natural to this audience.

Furthermore, we tested different versions of the two message variations we sent the most (Main and Equal Decision). The messages we sent were brief by design in effort to use less of the decision maker's time and to make our treatments about parent availability more salient. We did test longer versions of our two most-emailed messages, as detailed in Table J.1, but found that the difference in the callback rates was not statistically significant, nor was the proportion of calls to mothers versus fathers.

J.4 Pilot Studies

In May of 2021 we sent 767 emails, in June 2021 we sent out 1,250 emails, and in November 2021 we sent out 1,250 emails. The primary purpose of this early data collection was to refine the process by which we send emails, learn about response rates, and test our ability to

[12] https://namecensus.com/last-names/

[13] For example, if you Google "John List" versus "John List Economist," you will find that there is an infamous American murderer named John List who has fewer citations than the John List most economists know.

match phone calls to emails sent. As such, we concentrated on a subset of our treatments: Baseline, Male High Availability, Male Low Availability in the May and June 2021 waves, and expanded to five treatments in the November 2021 wave with the inclusion of the Female High Availability, Female Low Availability treatments.

Our pilot studies tested a number of procedural items. For our May pilot, we chose the names Jennifer and Michael because they signal gender well. However, Jennifer and Michael are predominantly white names, so we wanted to test a more race-neutral set of names (Erica/Roy) to see if this impacted callbacks. Testing Jennifer/Michael vs. Erica/Roy, we found that using the more race-neutral names (Erica/Roy) decreased callbacks by 8.8 percentage points. We felt that using the more race-neutral names increased the external validity of our findings and as such decided to use them in our full data-collection effort.

Additionally, we tested two types of email accounts in our pilots, given that our survey of educators indicated that the use of a joint family email address was less common than the use of individual email addresses and cc'ing the other parent (Section K). We found that using a joint family email address (versus individual email addresses, with one parent cc'ing the other parent) decreased our callback rates by 9.2 percentage points ($p = 0.032$). With the evidence from both the pilot and the survey, we dropped the joint family email address in our full data-collection efforts.

J.5 Phone Call Data

J.5.1 May 2022 Phone Calls

In May of 2022 we sent about eight thousand emails to schools, however, we found that some of these schools shared a single email address or a single phone number (e.g. a network of charter schools, or a school district that uses a centralemail address and/or central phone number). In addition, an error in our code meant we mistakenly sent more than one email to some email addresses. Removing all these from our dataset, we retained 7,935 emails sent to schools that each had a unique email and unique phone number.

In the weeks following, we received 2,990 callbacks to our May 2022 emails. ome of these callbacks are problematic: some are assumedly in response to emails we dropped from our dataset for the reasons outlined above,and a small number are likely spam calls made to our fictional parents' numbers (though these are most likely randomly distributed across our phone numbers). More of an issue is that these callbacks include calls made by the same school principal using multiple different phone numbers or just calling the same household multiple times in a row to the mother, the father, or some combination of both. Our outcome variable of interest is the first parent contacted, rather than the total number of calls made by a principal (although this could be of interest also). Furthermore, to be able to perform analysis about a school or principal's specific demographics, we need to link each phone call back to a specific email sent. This matching is a multistep process.

J.5.2 July 2022 Phone Calls

In July and August of 2022, we sent 72,136 emails. In the weeks following we received 30,214 calls. Much like our May data, these calls include spam calls. Our primary objectives with matching callbacks to specific schools is to allow analysis by attributes of the school and to identify correctly which parent was called first if calls were made from multiple phone numbers by the same school principal.

J.5.3 Matching Phone Calls To Emails

First we created a dataset with a single line for each unique phone number. We also included all the phone calls from "Restricted" phone numbers, as it is impossible to tell if those are unique. In May 2022 the one-call dataset had 1,684 lines, and in June/July 2022 the one-call dataset had 17,139 lines. We then matched these CallFire 10-digit phone numbers to the 10-digit phone numbers associated with our schools. A little over 60% of calls matched up.

We then took the remaining CallFire phone calls and performed a "fuzzy" match on the first 6 digits of each phone number. For example, all calls originating from Tufts University start with these same 6 digits, 617-627; all calls from Brigham Young University start with 801-422. We then had research assistants check these fuzzy matches for accuracy and disambiguation when two-plus schools matched to a single CallFire phone call. Around one-fifth of calls are matched by a "fuzzy" match.

For the remaining CallFire phone calls, we asked research assistants to listen to voicemails and perform Web searches to attempt to match them to a school we emailed.

Last, we randomly selected a subset of these matches to be audited by a different research assistant to check for the quality of our matching.

K Survey Evidence

In April 2022 we ran a survey of educators using Prolific (IRB number STUDY00002608). In April 2023 we ran a similar survey of adults who interact with parents, including educators. People were eligible to take our survey if they were over 18, reside in the US and regularly reach out to parents as part of their job. We had 238 educator respondents in 2022 and an additional 377 respondents from a variety of persons who interact with parents (the most common were Teacher, Childcare provider, Medical Practitioner, Nurse, Sports Leader). Of the 377 respondents in 2023, 77 self-identified as interacting with parents in the role of "other."

K.1 Educator Survey

In 2022, prior to fielding our experiment, one goal of our survey was to check that the type of email we were sending to schools was appropriate. Over 50% of educators reported getting the most questions about school enrollment during the month of August. August was followed by the months of May, September, July, June and April (in that order) with about 18% to 28% of educators stating they got the most questions about enrollment in these months. About three-fourths of educators said that being contacted by parents was either very common (at least once a week) or somewhat common (at least once a month). When being emailed by both parents a single parent emailing and cc'ing the other parent was more common than emails from a joint family email account. Educators reported they contacted parents by phone about the same amount as they did via email, email being slightly more common.

A second goal our survey was to see how educators self-reported calling mothers versus fathers in response to different types of inquiries. We found that educators self-reported they would make no call in response to a message like our main baseline only 8% of the time, this is very different than the rate we observe in our natural field experiment which is well above 80% not calling back either parent. This could be because some of email messages are going to spam, or because the group of survey respondents is a selected group, or because educators are overly confident in their likelihood of making a call. This disconnect highlights the importance of running a natural field experiment in this setting. Interestingly, conditional on self-reporting making a call the educators said they would call the female parent 57% of the time, which is quite similar to the rate we observe in the natural field experiment.

We found that educators always reported a higher level of wanting to contact the mother instead of the father if they had to choose a single parent to contact about a child being sick (98% contact mom), volunteering at a book fair (96%) or career day (78%), school related payments (86%), or a child's allergies (97%). We allowed the educators to rank the following reasons for choosing to contact the person which were displayed in a random order: I expect this person to be more likely to respond quickly, I expect this person to be more likely to be primary decision maker about this topic, I simply like interacting with this person more, and Other. The reasons of "I expect this person to be more likely to respond quickly", "I expect this person to be more likely to be primary decision maker about this topic" were very similarly ranked as the top choice within each type of inquiry.

K.2 Household Survey

Within our surveys we also identified which respondents were parents from a household with one male and one female parent. In April 2022 there were around 90 respondents who answered a series of questions about households and schools for us; in April 2023 just over an additional 125 parents answered questions about schools and other points of contact (e.g. Doctors, Law Enforcement, Sports). When asked "What proportion of the time does your child's school contact you versus your partner?" female parents report being contacted about 80% of the time while male parents are contacted about 40% of the time (note this sums to more than 100%, so each group may incorrectly perceive the reality of who the school contacts more). Mothers reported being contacted more than 50% of the time by doctors, sports leaders, extracurricular leaders, childcare providers, religious leaders, and other adults, while fathers reported less than 50% for all these decision-makers. The one place where mothers and fathers reported similar rates of being contact was by law enforcement.

Interestingly when asked how often they wish they were contacted by the child's school female parents report wanting to be contacted less, while male parents want to be contacted more. This same trend of women wanting to be contacted less and me wanting more of being happy with the current level was true for interactions with schools, sports leaders, extracurricular leaders, childcare providers, and other adults.

Our small sample which is not representative of the US as a whole has fairly similar results to a nationally representative survey which finds that in two-parent heterosexual households with school age children that mothers report being contacted 71% of the time, while males are contacted 48% of the time.[14] Mothers wish they were contacted less (65% ideally) while fathers in the national sample report wanting to be contacted almost exactly as much as they are contacted (at 47% of the time).

We also asked "When the school contacts your family, what proportion of the time do you respond first?" and found female parents report contacting the school first over 80% of time while male parents report making first contact only about 40% of time. For all the types of contacts mothers reported being more likely to respond first than fathers.

L Addressing Imbalance In Emails Sent From Mothers and Fathers

To address the unintentional imbalance in emails sent from mother's emails versus father's email in the main text we have weighted our observations so that emails from each parent are balanced. In this section we randomly delete observations from our data until we have achieved balance on emails from each parent, and find very similar results to those reported in the main text.

[14] https://csed.byu.edu/american-family-survey

Table L.1: **Alternative Sample Summary Statistics By Treatment in Main Variation Similar to Table 1**

	(1) High Male (Hm)	(2) Low Female (Lf)	(3) Baseline (b)	(4) Low Male (Lm)	(5) High Female (Hf)
FemaleNum0	0.05	0.10	0.12	0.15	0.19
MaleNum0	0.16	0.11	0.08	0.06	0.02
NoCall	0.79	0.79	0.80	0.79	0.78
FemaleNum	0.26	0.47	0.59	0.73	0.90
MaleNum	0.74	0.53	0.41	0.27	0.10
FemaleEmail	0.50	0.50	0.50	0.50	0.50
Observations	6728	5486	5016	5618	6066

Notes: FemaleNum0 is the proportion of calls made to a female parent when including NoCall as an outcome, and MaleNum0 is defined analogously. FemaleNum is a variable that takes a null value if no call was made, and thus represents the proportion of calls to female parents conditional on a call being made; MaleNum is defined analogously. FemaleMaleRatio can be computed either as the ratio of FemaleNum0 to MaleNum0 or as FemaleNum to MaleNum. FemaleEmail takes the value 1 if a the email was sent from the mother's email address and cced the father, and the value 0 if the email was sent from the father and cced the mother.

Table L.2: **Alternative Sample Summary Statistics By Variation (All Treatments Combined) Similar to Table 3**

	(1) Main	(2) Equal Decision	(3) Full Time	(4) Payments
FemaleNum0	0.122	0.121	0.117	0.100
MaleNum0	0.087	0.080	0.074	0.068
NoCall	0.791	0.799	0.808	0.832
FemaleNum	0.582	0.602	0.613	0.596
MaleNum	0.418	0.398	0.387	0.404
FemaleEmail	0.500	0.501	0.499	0.499
Observations	28914	28692	7983	8443

Notes: FemaleNum0 is the proportion of calls made to a female parent when including NoCall as an outcome, and MaleNum0 is defined analogously. FemaleNum is a variable that takes a null value if no call was made, and thus represents the proportion of calls to female parents conditional on a call being made; MaleNum is defined analogously. FemaleMaleRatio can be computed either as the ratio of FemaleNum0 to MaleNum0 or as FemaleNum to MaleNum. FemaleEmail takes the value 1 if a the email was sent from the mother's email address and cced the father, and the value 0 if the email was sent from the father and cced the mother.

Table L.3: **Alternative Sample Summary Statistics By Primary Email Sender Similar to Table 2**

Panel A: Email Sent By Mother (cc'ing Father)

	(1) All Messages	(2) High Male (Hm)	(3) Low Female (Lf)	(4) Baseline (b)	(5) Low Male (Lm)	(6) High Female (Hf)
FemaleNum0	0.17	0.08	0.18	0.20	0.21	0.20
MaleNum0	0.04	0.13	0.03	0.00	0.01	0.01
NoCall	0.79	0.78	0.79	0.80	0.78	0.79
FemaleNum	0.81	0.38	0.86	0.98	0.96	0.97
MaleNum	0.19	0.62	0.14	0.02	0.04	0.03
FemaleEmail	1.00	1.00	1.00	1.00	1.00	1.00
Observations	14448	3365	2726	2512	2813	3032

Panel B: Email Sent By Father (cc'ing Mother)

	(1) All Messages	(2) High Male (Hm)	(3) Low Female (Lf)	(4) Baseline (b)	(5) Low Male (Lm)	(6) High Female (Hf)
FemaleNum0	0.07	0.02	0.02	0.04	0.10	0.18
MaleNum0	0.13	0.18	0.19	0.16	0.11	0.04
NoCall	0.79	0.80	0.79	0.80	0.80	0.78
FemaleNum	0.35	0.12	0.08	0.21	0.48	0.83
MaleNum	0.65	0.88	0.92	0.79	0.52	0.17
FemaleEmail	0.00	0.00	0.00	0.00	0.00	0.00
Observations	14466	3363	2760	2504	2805	3034

Notes: FemaleNum0 is the proportion of calls made to a female parent when including NoCall as an outcome, and MaleNum0 is defined analogously. FemaleNum is a variable that takes a null value if no call was made, and thus represents the proportion of calls to female parents conditional on a call being made; MaleNum is defined analogously. FemaleMaleRatio can be computed either as the ratio of FemaleNum0 to MaleNum0 or as FemaleNum to MaleNum. FemaleEmail takes the value 1 if a the email was sent from the mother's email address and cced the father, and the value 0 if the email was sent from the father and cced the mother.

Table L.4: **Alternative Sample Multinomial Logit Models of Effect of Treatments on No Call, Call Male or Call Female Similar to Table A.1**

	(1)	(2)	(3)	(4)	(5)	(6)
	outcome	outcome	outcome	outcome	outcome	outcome
No_Call						
High Male (Hm)	-0.63***	-0.67***	0.80***	0.83***	0.00	0.00
	(0.06)	(0.06)	(0.07)	(0.07)	(.)	(.)
Low Female (Lf)	-0.28***	-0.29***	0.21**	0.22***	0.00	0.00
	(0.07)	(0.07)	(0.06)	(0.06)	(.)	(.)
Low Male (Lm)	0.37***	0.38***	-0.25***	-0.26***	0.00	0.00
	(0.08)	(0.08)	(0.06)	(0.06)	(.)	(.)
High Female (Hf)	1.30***	1.34***	-0.50***	-0.53***	0.00	0.00
	(0.10)	(0.10)	(0.05)	(0.06)	(.)	(.)
Female_Call						
High Male (Hm)	-1.43***	-1.50***	0.00	0.00	-0.80***	-0.83***
	(0.09)	(0.09)	(.)	(.)	(0.07)	(0.07)
Low Female (Lf)	-0.49***	-0.51***	0.00	0.00	-0.21**	-0.22***
	(0.09)	(0.09)	(.)	(.)	(0.06)	(0.06)
Low Male (Lm)	0.61***	0.64***	0.00	0.00	0.25***	0.26***
	(0.09)	(0.09)	(.)	(.)	(0.06)	(0.06)
High Female (Hf)	1.80***	1.87***	0.00	0.00	0.50***	0.53***
	(0.11)	(0.11)	(.)	(.)	(0.05)	(0.06)
Male_Call						
High Male (Hm)	0.00	0.00	1.43***	1.50***	0.63***	0.67***
	(.)	(.)	(0.09)	(0.09)	(0.06)	(0.06)
Low Female (Lf)	0.00	0.00	0.49***	0.51***	0.28***	0.29***
	(.)	(.)	(0.09)	(0.09)	(0.07)	(0.07)
Low Male (Lm)	0.00	0.00	-0.61***	-0.64***	-0.37***	-0.38***
	(.)	(.)	(0.09)	(0.09)	(0.08)	(0.08)
High Female (Hf)	0.00	0.00	-1.80***	-1.87***	-1.30***	-1.34***
	(.)	(.)	(0.11)	(0.11)	(0.10)	(0.10)
Control Variables						
Observations	28914	28914	28914	28914	28914	28914

Notes: This table presents the results of a multinomial logit model using a model like the one in Equation 10. The outcome variable takes three values: no call, call female, or call male. In this table we present the results with a base case of no call in columns (1) and (2), female call in columns (3) and (4), and male call in columns (5) and (6). The results from the three base cases are analogous and all three are presented to make specific comparisons more simple. The outcomes with no controls from this table are represented visually in Figure B.1.

Table L.5: **Alternative Sample Multinomial Logit Models For Theory Model Similar to Table A.2**

	(1) Main	(2) Expertise	(3) Payment	(4) Full Time
Female_Call				
any_msg_M	-0.32***	-0.20***	-0.35***	-0.28*
	(0.05)	(0.05)	(0.11)	(0.11)
x_M	-0.51***	-0.35***	-0.54***	-0.44***
	(0.03)	(0.04)	(0.07)	(0.07)
any_msg_F	0.09+	0.31***	0.04	0.37***
	(0.05)	(0.05)	(0.10)	(0.10)
x_F	0.36***	0.35***	0.36***	0.28***
	(0.03)	(0.03)	(0.06)	(0.05)
Constant	-1.85***	-2.04***	-2.09***	-2.11***
	(0.04)	(0.04)	(0.08)	(0.09)
Male_Call				
any_msg_M	0.12*	0.21***	0.26*	0.32**
	(0.06)	(0.06)	(0.12)	(0.12)
x_M	0.50***	0.41***	0.49***	0.40***
	(0.03)	(0.03)	(0.07)	(0.06)
any_msg_F	-0.53***	-0.35***	-0.39*	-0.54***
	(0.07)	(0.06)	(0.15)	(0.14)
x_F	-0.78***	-0.50***	-0.82***	-0.75***
	(0.05)	(0.04)	(0.11)	(0.10)
Constant	-2.25***	-2.33***	-2.67***	-2.45***
	(0.05)	(0.05)	(0.10)	(0.10)
N	28914.00	28692.00	8443.00	7983.00

Notes: This table presents the results of a multinomial logit model using a model like the one in Equation 10. The outcome variable takes three values: no call, call female, or call male. In this table we present the results with a base case of no call. The right hand side variables are discussed in Section 2.

Table L.6: **Alternative Sample Balance on Observable Attributes of Schools/Decision Makers By Treatment In Main Variation Similar to Table C.1**

	(1) High Male (Hm)	(2) Low Female (Lf)	(3) Baseline (b)	(4) Low Male (Lm)	(5) High Female (Hf)
Elementary	0.48	0.49	0.51	0.50	0.50
Middle	0.14	0.14	0.14	0.15	0.15
High	0.19	0.20	0.20	0.19	0.20
Decison-Maker Female	0.57	0.58	0.59	0.59	0.58
PublicCharter	0.06	0.05	0.06	0.06	0.06
PublicNOTCharter	0.76	0.79	0.81	0.79	0.80
Private	0.18	0.16	0.13	0.15	0.14
FreeLunch	0.55	0.56	0.54	0.55	0.53
White	0.53	0.52	0.52	0.53	0.52
Black	0.14	0.15	0.14	0.14	0.15
Hispanic	0.23	0.22	0.23	0.23	0.23
FemaleEmail	0.50	0.50	0.50	0.50	0.50
Observations	6728	5486	5016	5618	6066

Notes: There is a small proportion of schools which are not Elementary, Middle or High Schools (e.g. K-12 or pre-schools). The following variables are only known for non-private schools: FreeLunch, White, Black, Hispanic. DMFemale is whether the decision maker (the principal) has a first name that is female.