WHO IS SEX-SELECTING, AND WHEN?

Minu Philip*

October 2024

Abstract

Parents with a strong preference for sons are known to engineer the sex-composition of their children. They do so either by selectively aborting female fetuses (sex-selection) or by continued childbearing until a son is born. I propose and empirically validate a general heuristic or rule that describes *when* parents decide to sex-select. Using data on mothers' birth history and their self-reported ideal number of children, I define *relative birth orders* for each child that indicates how far their birth order is from the mother's ideal number of children. I find the ratio of male-to-female births to be the highest for children born at their mother's ideal number of children. This suggests a heuristic whereby parents sex-select when at their ideal number of children, to avoid exceeding it. I empirically validate this heuristic by exploiting the natural orthogonality between sex assigned at birth and the birth interval that precedes it. Following birth histories with no or few male births, I find intervals preceding male births to be longer for parents who are at their ideal parity than others—as the heuristic suggests. I also find evidence of reversals in the decision to sex-select following unsuccessful attempts, and of the use of sex-selection even at the first birth order.

*Department of Economics, New York University.
I am grateful to Professors Debraj Ray and Martin Rotemberg for their continuous guidance and counsel that has refined this project to its current form. This project has also benefitted from conversations with Alberto Bisin, Katarína Borovičková, Guillaume Fréchette, and Prashant Bharadwaj.

I. INTRODUCTION

Son-preference refers to a widespread preferential bias for male children. Motivated by such preferences, parents are known to engineer the sex-composition of their children. This is known as 'son-targeting', and usually entails the use of one of two methods: *differential stopping behavior*, referring to continued childbearing until a desired number of sons are born; and *sex-selection*, which is the selective abortion of female fetuses in-utero. These practices skew birth sex-ratios, fertility rates, and the order and spacing of children; in turn having far-reaching implications for broader aspects such as population demographics, marriage market equilibria, maternal well-being, and child health. The use of differential stopping and sex-selection has been evidenced extensively in the literature, with the consequences also discussed therewith. What is not known, is *when* parents decide to sex-select instead of continuing to rely on the natural odds of a male birth.

In this paper, I identify a general heuristic or rule that describes when son-preferring parents decide to sex-select. I analyze birth histories of Indian women as recorded in the National Family Health Surveys (NFHS) to infer the son-targeting practices that parents resort to and the precise heuristics that they follow.[1] India presents as a suitable context for this study given the widespread prevalence of son-preference as suggested by its very high sex-ratios. Moreover, India's ongoing fertility transition generates sufficient variation in fertility rates to facilitate our analyses, unlike other low or high-fertility countries.

Identifying a heuristic of parental manipulation is, however, quite challenging. This is so because parental manipulations are neither directly observed nor reported, and the preferences that motivate these choices are also typically not observed.[2] Any attempt at identifying such a heuristic must necessarily include an inference of fertility preferences and decisions from realized birth histories. That said, even with detailed birth histories, it is not trivial to separate instances of sex-selection from differential fertility stopping. This is why the literature has often studied the two separately.[3] Importantly, the literature has studied the use of these two methods in the aggregate, with the specific aspects of what determines parents' choice between the two methods, or when they switch between either, largely being unexplored.

While both the methods are motivated by a desire for more sons, sex-selection in particular is necessitated by a high cost of continued childbearing. I rely on this distinguishing parental motivation to identify when parents decide to sex-select. More specifically, I define a *relative birth order* for each child indicating the distance between its birth and the mother's self-reported ideal number of children, or *ideal parity*. For example, a mother with an ideal parity of two, who gives birth to three children (at birth orders 1, 2 and 3) is described to have given birth at corresponding relative orders -1, 0 and 1; with the last birth being *one more than* what the mother had desired.

[1] NFHS is the India-specific implementation of the Demographic and Health Survey (DHS) framework.

[2] The use of pre-natal sex determination of fetuses is illegal in India under the Pre-conception and Prenatal Diagnostic Techniques Act (PCPNDT, 1994). Parents also misrepresent their preference for sex-composition to avoid acknowledging their preferential bias for sons.

[3] The exception to this is Baland et al. (2023), that studies the two methods under a unified framework and proposes tests to detect the use of differential stopping and sex-selection at various birth orders.

Relative birth orders represent the increasing cost of successive birth at each birth order. Additionally, they also allow us to compare births at different birth orders for mothers who have different ideal parities. That is, it reflects that a child born at the second birth order means differently to a mother with an ideal parity of two than it does to mother with an ideal parity of four. At the same time, this child is also somewhat comparable, in terms of parental fertility incentives, to a child born at the fourth birth order to a mother with an ideal parity of four.

Comparing the proportion of sons at various relative birth orders, I find significantly high proportion of male births when mothers are at their ideal parity. Since parity preferences should theoretically be orthogonal to outcomes such as sex or birth history in the absence of parental manipulations, high proportion of male births when mothers are at their ideal parity strongly implies the use of sex-selection at the ideal parity. This suggests a heuristic whereby son-preferring parents who want to sex-select typically doing so when at their ideal parity, in order to avoid exceeding it.

For a more rigorous examination of this deduced heuristic, I exploit the natural sex- determination system in humans that renders the sex of a newborn independent of the sex-composition of any children already born to the mother. In India, where the revelation of the sex of the fetus in-utero is illegal, sex of a child should be revealed naturally *and legally* only at the time of birth, i.e. at *end* of the *preceding* birth interval. The length of the preceding inter-birth interval and the sex of newborn born at the end of it, should therefore be uncorrelated and independent of the sex-composition of children already born to the mother. The use of sex-selection, however, artificially increases the *observed* birth interval preceding a live male birth. This increase is on account of the inability to determine sex of a fetus before at least 10-12 weeks of pregnancy, the time needed for post-abortion recovery, and the time taken till the next conception (Pörtner, 2016, 2022).[4] Even if there were other reasons that rendered intervals preceding male births to be longer, they should be common to all birth histories. Any significant difference in the length of the preceding interval based on the sex of the newborn *across* histories must then be attributed to parental manipulation. This is because the incentives to sex-select vary with birth history, while the effect of other confounding variables do not.

To identify use of sex-selection, I test whether sex of a newborn "affects" the length of the interval preceding its birth. For this, I use a difference-in difference specification that compares the length of the birth interval preceding male and female births across different birth histories. The identifying assumption here—akin to the 'Parallel Trends Assumption'—is that in the absence of sex-selection, the difference in birth-intervals preceding male and female births shouldn't be different across birth histories. This "assumption", to our convenience, is guaranteed by human biology as sex is determined independently of previous births.

To validate the heuristic, I use the difference-in-difference specification to test for sex-selection among sub-samples of parents that the heuristic suggests would sex-select versus those who wouldn't. I find that any birth order, sex-selection is driven by the mothers who

[4]Even in the case of pre-implantation sex-selection methods (such as embryo sorting) that use assisted reproductive technologies, it can take several cycles until the curated embryos are successfully implanted.

report an ideal parity equal to that birth order. Extending the analysis to the first birth order, I also find evidence of mothers with an ideal parity of one resorting to sex-selection as early as at the first birth order.

I find these empirical patterns to be consistent across three rounds of NFHS: Round 3 surveying 67,451 women who gave birth between the years 1978 and 2005; Round 4 surveying 366,224 women who gave birth between 1987 and 2015; and the Round 5 surveying 375,004 women who gave birth between 1991 and 2019. These patterns are also similar in rural and urban areas, and across levels of maternal education. Though the ideal parity of mothers varies with their level of education, the heuristic they follow to sex-select does not. Heterogeneity is observed only along the lines of geographical region, that is synonymous with cultural divisions in India. North-western states such as Rajasthan, Punjab and Haryana as well as states in Western and Central India, display strong empirical evidence of sex-selection and at the ideal parity, and consistently over the three rounds. The Southern and North-Eastern states, in contrast, display no such trend.

My findings provide insight into the heuristic followed by parents who have to continually choose between heeding to the natural odds to birth orders as opposed to engaging in sex-selection. More importantly, in light of the continuous decline in desired parity, these findings warn against any simplistic expectations about the use of sex-selective methods declining with economic development or increased education. With a negative correlation between desired parity and both economic development and the level of education, the problem of sex-selection would only become more pertinent.

The remainder of this paper is organized as follows. Section II offers a background on son-prevalence and the data indications of son-targeting methods. It also details the data used in my analysis, and examines the prevalence of the two methods of son-targeting methods. In Section III, I detail the conceptual framework. First, Section III.B describes the empirical patterns parents' fertility choices would generate for observed birth-statistics such as proportions of sons and inter-birth intervals if parents adopted specific son-targeting methods. This discussion guides the inference of the heuristic. Next, Section III.C details the specification used to empirically validate the heuristic. Section IV applies this to data to infer the heuristic and validate it. Finally, Section V concludes.

II. CONTEXT AND DATA

II.A. Son-targeting methods and their indications

A preference for sons over daughters is globally widespread, especially in South and East Asia and in North and East Africa. It generally stems from a combination of cultural, religious and economic reasons. In the case of India, it originates from patrilocal kinship and marriage norms that precipitate sentiments of girl children being non-members at their natal homes ("*paraya dhan*") and effectively a liability ("*bojh*") until they're married and sent away. Additionally, with men being typical earners in any household and also recipients of dowry payments in many cultures, sons are viewed as providers of financial support to parents in their old age; furthering the preferential bias. Religious rites and inheritance laws also grant sons with important roles, leaving parents wanting a son to carry forward their familial name, lineage, and wealth. The geographical prevalence of son preference in India, therefore, mirrors the geographical divide in religious and cultural practices. Son-preference is much more prevalent in northern and western states of India than in the historically matrilineal south or north-east. Pande & Astone (2007) and Jayachandran (2015) describe the relevance of various such socioeconomic and socio-cultural factors in explaining son preference both at the individual and community level.

Motivated by son-preference, parents engineer the sex-composition of their children by resorting to differential fertility stopping or sex-selection. The use of these son-targeting methods by parents in India has been well documented in the literature based on their respective empirical indications, as outlined below.

Differential stopping is indicated by high parity progression ratios (i.e. the proportion of women who go on to have another child) for birth histories with fewer sons, as compared to other histories at the same parity (Das, 1987). It results in girl children having more siblings on average, and most likely being the older sibling in these large families (Basu & Jong, 2010). Baland et al. (2020) formalize these effects of differential stopping to identify countries where stopping rules are used: these are countries where at any birth order, girls have more younger siblings than boys.Differential stopping also affects birth spacing or the time interval between two consecutive births. Parents with an unfulfilled demand for sons tend to begin attempts at conception earlier than others, resulting in shorter birth intervals succeeding female births (Javed & Mughal, 2020). This translates to shorter breastfeeding duration for daughters born at lower birth orders (Jayachandran & Kuziemko, 2011). Consequently, girl children of parents engaging in differential stopping are subject to higher sibling rivalry and competition for nutritional investments (Jayachandran & Pande, 2017).

The other method of son-targeting, sex-selection, is where parents selectively eliminate unwanted female fetuses in-utero. Its use is neither reported or openly admitted to because it is unlawful under the Prenatal Sex Diagnostic Techniques Regulation and Prevention of Misuse (PNDT) Act of 1994. The Act criminalizes misuse of prenatal sex-diagnostic techniques to reveal the sex of a fetus, effectively banning prenatal screening of sex. The ban has nevertheless been hard to enforce and proved ineffective. Bhalotra & Cochrane (2010) estimate as many as 0.48 million girls to have been selectively aborted each year between 1995-2005. Later amendments to the Act (in 2002) also haven't improved the situation.

Sex-selection is indicated by skewed sex-ratios, expressed as the number of females per 1000 males (Sen, 1990, 1992; Jha et al., 2006); and also results in the proportion of girls among elder siblings to be larger for a boy than for a girl (Baland et al., 2023). Its use is also indicated by longer inter-birth intervals preceding male births relative to female births (Pörtner, 2016, 2022). This is because sex can be determined only after 12 weeks of pregnancy, and the time taken for the next male conception and gestation add up to what is observed as a long inter-birth interval preceding male births.

The literature has been mostly silent about whether parents exclusively choose one of the two methods, or if they switch between the two. And if they decide to sex-select, either with or without differential stopping, when do they resort to it?

II.B. Data

To study parents' choice of sex-selection over differential stopping and how it relates to their fertility preferences, I examine retrospective birth histories of Indian women aged between 15-49 years, as recorded in the National Family Health Survey (NFHS). The NFHS surveys collect information on female respondents' birth history, fertility preferences and sexual health practices. I use data from three rounds of the NFHS surveys: Round 3 (2005-06), Round 4 (2015-16) and Round 5 (2019-21). The sample from Round 3 surveys 67,451 women bearing children between the years 1978 and 2005; the one from Round 4 surveys 366,224 women bearing children between the years 1987 and 2015; and the one from Round 5 surveys 375,004 women bearing children between the years 1991 and 2019.[5] For each round, I restrict the sample to women who have had at least one birth and who report having completed their fertility.[6] I exclude mothers who have had multiple births in a single pregnancy. Effectively, with each round, I examine birth histories of women who finished childbearing in different decades.

Respondents report the details of each birth, including the length of the inter-birth interval, the birth order and sex.[7] The surveys also collects relevant information on the respondents, relating their age, level of education, household characteristics, wealth index,[8] contraceptive use, reproductive health as well as their fertility preferences in the form of ideal parity and ideal sex-composition. More specifically, respondents are asked to report their ideal number and the ideal composition of children—number of sons, daughters and those of either sex—that *"the respondent would like to have in her whole life, irrespective of the number she already has."*

Table 1 summarizes the fertility preferences and outcomes of Indian women over the three rounds. While women's age at the beginning and end of childbearing has largely remained the same over the years, there has been in steep decline in actual parity. This decline cannot be attributed to a commensurate change in women's preferred or ideal parity, which has instead remained fairly stable. It must then be that either parents are able to achieve their desired composition at earlier parities, or have increasingly become unwilling to trade-off low parity for a more desired sex-composition.

An important concern that may arise is over the reliability of reported ideal parity, and how must it be rightly interpreted. They may, for example, may be ex-post rationalized based on the actual or 'realized parity. As a preliminary test, I confirm that the correlation between actual and ideal parity isn't high. Moreover, this correlation is even lower for respondents

[5] The range of birth years is based on the 5th and 95th percentiles respectively.

[6] Respondents are asked if they would like to have another child, to which they reply if they would, wouldn't, haven't decided, or cannot because they are either sterilized or declared infecund. All women who report not wanting any more children, or being unable to bear any more children are classified as having *completed* their fertility.

[7] For all births reported, the survey confirms whether the child is alive at the time of the survey and if not, the age at death. In case the respondent fails to report births that were subject to infanticide, it would be difficult to separate instances of unreported female infanticide from longer birth intervals.

[8] The wealth index is a composite measure of the household's living standard based on aspects such as the household's ownership of assets, access to living amenities, and type of housing, among others. It is defined on a continuous scale and is used in the data to categorize households into quintiles of relative wealth.

Table 1 :
Descriptive Statistics

	NFHS Rounds		
	3 2005-06	4 2015-16	5 2019-21
Median age at first birth	20	20	20
Median age at last birth	26	26	25
Parity (Mean)	3.278	3.016	2.796
Proportion of sons (Mean)	0.551	0.565	0.563
Ideal parity (Mean)	2.399	2.399	2.322
Ideal number of sons (Mean)	1.136	1.225	1.150
Ever used modern contraception (excludes permanent methods)	0.393	0.242	0.348
No. of respondents	67,451	366,224	375,004

Notes: This table summarizes the fertility preferences and realizations of mothers who report having completed their fertility. The data is drawn from the 2005-06, 2015-16, and 2019-21 rounds of the National Family Health Survey for India.

who have had at least one instance of a multiple birth (Appendix Table B1), indicating that reported 'ideal' parity doesn't exactly tail actual parity.[9]

I interpret the ideal parity figures as the preference that respondents in my sample based their fertility choices on. This interpretation warrants some clarifications. First, respondents' preferences for ideal parity and composition could potentially vary over time with their birth experiences and history. Considering that respondents in the sample have completed their fertility, I interpret reported preferences as those that would have been relevant for fertility choices made at the time of their last birth and hence relevant to the decision of completing fertility or stopping childbearing. Second, preferences that guide fertility choices are often not of the mother alone, but also influenced by the preferences of her partner and other family members. Once again, considering that respondents in the sample have completed their fertility, it is reasonable to expect them to be aware of their partner's fertility preferences. More than 75% of women in the 2005-06 sample believe that their partner "wants the *same* number of children[...]that she wants herself." This statistic increases to about 85% in the 2015-16 and 2019-21 samples. Whether this alignment in fertility preferences between the two partners exists already, or occurs after bargaining does not matter since we only concern ourselves with the need to capture preferences that ultimately influence the fertility decisions.

[9] Jayachandran & Kuziemko (2011) also use reported ideal parity in their analyses and discuss some concerns that it may raise.

II.C. Prevalence of son-targeting: Crude estimates

An important aspect that underscores the relevance of identifying heuristics of sex-selection, is that its use is still prevalent. This section undertakes an analysis of how prevalent the two son-targeting methods are and whether there has been any change in their use. Although the use of sex-selection or differential stopping is not directly reported or admitted to, we can deduce crude proportions of how many parents use either methods, under some assumptions. In this exercise, I follow the following definitions of the two son-targeting methods:

Let h represent a birth history i.e. a unique sex-sequence of children already born to a parent.

Definition 1 *: **Differential stopping behavior***
If $\exists$ a pair of birth histories $h \neq h'$, with equal number of children, say n, but dissimilar sex-composition, such that the tendency to stop childbearing following h differs from that following h', then it is indicative of differential stopping behavior at parity n.

Definition 2 *: **Sex-selection***
If any positive measure of parents selectively eliminate female births given history h, it is indicative of sex-selection following history h.

Let us suppose that the natural probability of giving birth to a male child were known, and identical for all mothers (say, π). Also assume that all sex-selecting mothers abort female fetuses only up to T times (if necessary, that is) to conceive a male child. After T failed attempts, they give up and stop conceiving.

Then, if

$$\begin{aligned} N_h &: \text{number of mothers whose birth history begins with the sequence } h \\ \mu_h &: \text{proportion of mothers who stop childbearing following history } h \\ \lambda_h &: \text{proportion of mothers who sex-selectively abort following history } h; \end{aligned}$$

following any history h,

$$\begin{aligned} \pi(1-\mu_h)N_h &\text{ mothers would birth a successive male child naturally} \\ (1-\pi)(1-\lambda_h)(1-\mu_h)N_h &\text{ mothers would birth a successive female child} \\ \lambda_h(1-\pi)(1-\mu_h)N_h &\text{ mothers would sex-select} \end{aligned}$$

Of the $\lambda_h(1-\pi)(1-\mu_h)N_h$ mothers who would selectively abort, a proportion π would birth male children in their second attempt, followed by a proportion $\pi(1-\pi)$ who would birth a male child in their third attempt and so on, up to $T < \infty$ attempts.

With no take-back or reversal of the decision to sex-select, total male births at the successive order will be

$$\underbrace{\pi(1-\mu_h)N_h}_{\text{male birth at first attempt}} \quad + \underbrace{\lambda_h(1-\pi)(1-\mu_h)N_h\left[\pi + \pi(1-\pi) + \pi(1-\pi)^2 + ... + \pi(1-\pi)^{(T-1)}\right]}_{\text{male births after selective abortions}}$$

$$= \left\{ \pi + \lambda_h (1-\pi)[1-(1-\pi)^T] \right\} (1-\mu_h) N_h$$

Total successive births will be

$$\left[\underbrace{(1-\pi)(1-\lambda_h)}_{\text{female}} + \underbrace{\pi + \lambda_h(1-\pi)[1-(1-\pi)^T]}_{\text{male}} \right] (1-\mu_h) N_h$$

The proportion of female successive births is hence given by

$$\frac{\left[(1-\pi)(1-\lambda_h)\right]}{\left[1-\lambda_h(1-\pi)^{T+1}\right]} \tag{1}$$

Let $\hat{p_h}$ be the proportion of male successive births observed in the sample following history h. Equating $1-\hat{p_h}$ to the theoretical proportion of female successive births, we can obtain a crude *estimator for the proportion of sex-selectors* at history h, given by

$$\hat{\lambda_h} = \frac{(1-\pi)-(1-\hat{p_h})}{(1-\pi)-(1-\hat{p_h})(1-\pi)^{T+1}} = \frac{\hat{p_h}-\pi}{(1-\pi)-(1-\hat{p_h})(1-\pi)^{T+1}} \tag{2}$$

It is the ratio of excess male births scaled by the probability of male birth within T sex-selection attempts by parents whose first conception after history h is of a female fetus. The higher the number of attempts parents are willing to make, the lower is the proportion of sex-selectors given any sample.

With the proportion of sex-selectors known, we can also back out the proportion of parents who decide to stop childbearing following history h i.e. the fertility stoppers. Had all non-sex-selecting parents following history h conceived again, there would have been $(1-\pi)(1-\lambda_h)N_h$ successive female births. The crude proportion of fertility stoppers after h can therefore be backed out by the percentage shortfall in the number of actual successive girls born from the number of girls who *would have been born* had all non sex-selecting parents had another child.

$$\hat{\mu_h} = 1 - \frac{N_{h,\mathtt{F}}}{(1-\pi)N_h(1-\hat{\lambda_h})} \tag{3}$$

Table 2 reports crude estimates of the proportion of fertility stoppers and sex-selectors based on births recorded in three rounds of the data. Positive proportions of sex-selectors is indicative of use of sex-selection; and *differences* in the proportion of fertility stoppers at the same parity but for histories with different sex composition, is indicative of differential stopping behavior. Since live births are known to consistently be male-biased in nature, $\pi = 0.512$ is used to generate the estimates based on Orzack et al. (2015). Table 2 also assumes the maximum number of selection attempts to be $T = 3$; lower values would increase the estimated proportion of sex-selectors and values of $T > 3$ wouldn't lower the estimates by a lot. It is important to acknowledge that there is no consensus on what the appropriate values of π and T should be, and whether they should be the same for all parents or not.

Nevertheless, this assumption is made for only exploratory comparisons. Although proportions are ideally supposed to be between 0 and 1, the estimates in Table 2—based on crude calculations—sometimes take negative values when they are close to zero. The objective here is not to take the estimates at face value, but to make relative comparisons across histories and across time periods to deduce ballpark proportions of parents who manipulate or engineer the sex-composition of their children.

TABLE 2 :
Crude estimates of sex-selectors and fertility stoppers

	Proportion of		Proportion of		Proportion of	
	Stoppers	*Sex-selectors*	*Stoppers*	*Sex-selectors*	*Stoppers*	*Sex-selectors*
NFHS Round	2004-05		2015-16		2019-21	
-	N/A	0.04	N/A	0.07	N/A	0.06
`M`	0.10	0.01	0.11	0.01	0.12	-0.01
`F`	0.07	0.10	0.06	0.11	0.07	0.11
`MM`	0.39	0.00	0.48	-0.02	0.54	-0.05
`MF`	0.36	0.04	0.42	0.04	0.49	0.02
`FM`	0.37	0.04	0.44	0.05	0.51	0.05
`FF`	0.19	0.12	0.19	0.15	0.23	0.17
`MMM`	0.43	-0.02	0.50	-0.02	0.57	-0.05
`MMF`	0.47	0.00	0.55	-0.02	0.62	-0.02
`MFM`	0.50	0.02	0.59	0.00	0.66	-0.04
`FMM`	0.50	-0.01	0.60	0.00	0.67	-0.02
`FFM`	0.44	0.03	0.50	0.05	0.58	0.09
`FMF`	0.35	0.06	0.40	0.05	0.46	0.07
`MFF`	0.34	0.08	0.39	0.06	0.46	0.06
`FFF`	0.21	0.11	0.21	0.15	0.25	0.15
Observations	67451		366224		375004	

Notes: This table reports crude estimates of the proportion of sex-selectors and fertility stoppers following select birth histories, using Equations (2) and (3). In the calculation of these numbers, I assume: (1) π, the natural probability of conceiving male child = 0.512, and (2) T, the maximum number of abortions that sex-selectors are willing to undergo = 3. Since these estimates are based on crude estimators, select values close to 0 are negative.

Table 2 suggests that for all histories, the proportion of fertility stoppers has increased over time, as can be reconciled with the decline in total parity over time. This increase in fertility stoppers has been the most modest for histories with only female births i.e. `F, FF, FF` where the proportion of sex-selectors has increased with time—indicating the continued prevalence of differential stopping.

Likewise, the use of sex-selection is neither inconsequential nor has it declined over the years. Consider the history `F`, roughly 93% of parents with a history of `F` go on to birth a second child, of which about 11% resort to sex-selection to secure a male child in particular. Note that the 11% who resorted to sex-selection did so because they conceived a female fetus. The proportion of parents who would have been *willing* to sex-select if necessary, unbeknownst of the sex of the fetus that they would conceive eventually, is probably much higher. The estimates of sex-selectors reported in Table 2 are also not too far from previously reported estimates of 11.6% of births at the second birth order and 15.3% of births at the third birth-order being subject to prenatal sex diagnosis between 1995-2005 as reported in Bhalotra & Cochrane (2010).

Contrary to assumptions made previously in the literature, the estimates also suggest that parents may be sex-selecting even at the first birth order. We confirm this observation more rigorously in Section IV.C.

III. CONCEPTUAL FRAMEWORK

This section begins with a general framework that describes parents' fertility decisions and the nature of data that it generates. Section III.B builds on this framework to detail the empirical patterns that specific son-targeting methods would generate for observed birth-statistics such as proportions of sons and inter-birth intervals; providing intuition for the identification of parental heuristics. I apply this intuition later in Section IV to infer the heuristic of when parents decide to sex-select. Section III.C then details the empirical strategy I use to validate the identified heuristic.

III.A. Setup

I model parents to be unitary non-gendered agents who, given their fertility preferences and concurrent birth history, decide whether to conceive again, and whether to sex-select upon conception.[10] Any parent with birth history h could stop childbearing and remain at h or could transit to either $(h, \texttt{F})$ or $(h, \texttt{M})$. Sex-selection is a choice that parents can exercise to limit the possible paths at any history to either remaining at h (by selective abortion) or transiting only to $(h, \texttt{M})$.

Parents begin with a null history or an empty sequence, and based on their actions and birth realizations settle over time into various 'nodes' representing their birth histories—capturing the order and sex of each child born, as well as the sex-composition up to each birth order.

Preferences

Parents' preferences are taken to be a complete and transitive ordering over all possible sex-sequences including an empty sequence representing no births. At the minimum, this ordering is based on parity i.e. the number of elements in the sex-sequence. For several other parents, son-preferring or not, it would also depend on the sex-composition.

Let n^* be the ideal parity for a parent, defined as the total number of children in the most-preferred sex-sequence of the parent. With an assumption that parents monotonically prefer

[10] I abstain from modeling any gendered differences in fertility preferences or any bargaining between spouses (see for example, Rasul (2008)).

lower parities (all else equal), n^* is well-defined and unique for each parent.[11] Likewise, the 'ideal' number of sons or daughters would be the respective frequencies of `M` or `F` in their most preferred sex sequence.[12] Let b^* be the ideal number of sons for the parent, defined as the number of boys in the maximal sex-sequence of the parent's preference ranking.[13]

Examples: For a parent who prefers a son and is otherwise indifferent over parity, based on the definitions, $n^* = b^* = 1$. For a parent who wants equal number of boys and girls, $n^* = 2, b^* = 1$. For a parent who wants two children and is indifferent about the composition, $n^* = 2$ and b^* is undefined.

Key Definitions

At each birth, son-preferring parents are faced with the choice of either playing these natural odds of a male birth, or manipulating the final composition of their children by sex-selecting. While playing the natural runs the risk of birthing an additional child of an unwanted sex, sex-selection is also a costly alternative (where the costs include illegality and hopefully morality). The decision to sex-select, therefore, effectively comes from a desire for a more favorable sex-composition but is necessitated by a high cost of additional births.

To capture this trade-off between the gains from a more desirable composition relative to the cost of an additional birth, I define *relative birth orders* that indicate how far each birth is from the respective mother's ideal parity.

Definition 3 ***Relative birth order with respect to ideal parity***
Let k denote the birth order of a child and n^* denote the ideal parity of the parent; then the relative birth order with respect to ideal parity for the child is defined as $j^* = k - n^*$. Note that $j^* > 0$ for all births that exceed the parent's ideal parity.

Example: Consider a parent who desires two children, but birthed three; i.e $n^* = 2,$ and $N = 3$. Her third child, born at order $k = 3$, is the youngest child but is one over the ideal. That is, the child is born at $j^* = 1$. For a parent who desires and births exactly the same number of children, her youngest child is born at $j^* = 0$ and is well within her ideal parity.

Relative birth orders allow the comparison of births at different birth orders for mothers who have varying parity preferences. It reflects that a child born at the second birth order means differently to a mother with an ideal parity of two (i.e. $j^* = 0$) than it does to mother with an ideal parity of four (i.e. $j^* = -2$). At the same time, this child is also somewhat comparable, in terms of parental fertility incentives, to a child born at the fourth birth order to a mother with an ideal parity of four with $j^* = 4$.

As j^* increases, the cost of an additional birth increases. For $j^* \leq 0$, there is net benefit from each additional birth; but $j^* \geq 1$ onwards, each additional birth has a positive cost potentially increasing in j^*. Parents would give birth at $j^* > 0$ only if the expected gains from a more desirable composition were high enough to offset the cost of the additional birth.

[11] If P^* is the set of most preferred or maximal combinations of (n, b), then $n^* = \min\{n | (n, b) \in P^*\}$.

[12] Note that here we are assuming that parents' *preferences* are not dependant on the specific ordering within a sex sequence.

[13] For $B = \{b | (n^*, b) \in P^*\}$. I take $b^* \in B$ iff $|B| = 1$, otherwise, b^* is undefined for parents who are indifferent about the sex-composition at n^*.

III.B. Heuristics of Sex-Selection

Sex-selection at any birth order k is typically indicated by a skewed ratio of male-to-female births at the order. The use of differential stopping, on the hand, does not skew birth sex-ratios *in the aggregate*. This distinction between the two son-targeting methods, however, breaks down for disaggregated data. For example, suppose we look exclusively at last births: Differential stopping could also produce high male-to-female birth ratios at the last birth because the decision to stop childbearing is made only when a son (or some desired number of sons) is born. Similarly, if we disaggregate by a parent's parity, differential stopping is also sufficient to produce high male-to-female birth ratios at the lower birth orders (Clark, 2000). This is because it is those parents with few or no sons who are selectively more likely to birth more children and have a higher parity. There is, therefore, a selection of parents who give birth to daughters at lower orders *into* higher parity.

Since the inference of any parental heuristic requires that parents' preferences be accounted for, it calls for disaggregation. Here, examining sex-ratios at *relative birth orders* i.e. at j^*s has the advantage that it both accounts for heterogeneity in parents' ideal parity preferences and does not suffer from the congruence in sex-ratio patterns generated by the two son-targeting methods. The intuition here is that sex realizations at birth are independent of parental preferences, and therefore independent of n^* and thereby also of j^*. In the absence of sex-selection, the ratio of male-to-female births is realized exogenously to n^* and therefore must equal the natural sex-ratio at all j^* no matter the dimension of disaggregation. Skewed sex ratios at j^* can therefore be unambiguously attributed to the use of sex-selection.

Moreover, comparing births at the same *relative* birth order ensures comparison between births when parents are faced with similar 'costs' of an additional birth. Though the cost of an additional birth weakly increases with each birth order for all, the increase may be larger for some parents with a lower value of ideal parity than for others. This affects their fertility decisions and their decision to sex-select, in particular.

Therefore, to infer heuristics of *when* parents choose to sex-select, I examine sex-ratios at all *relative birth orders.*

Proposition: *The use of sex-selection can be unambiguously inferred from skewed sex-ratios at relative birth orders (j^*).*

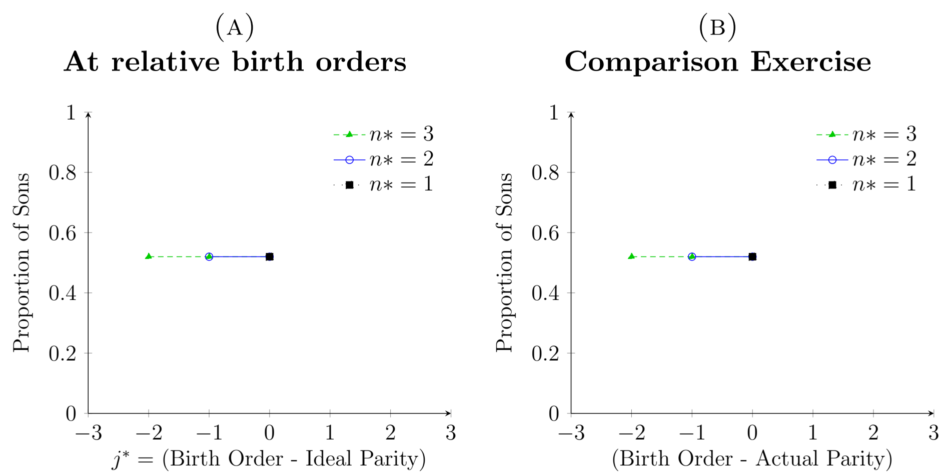
1. *In the absence of any manipulation, parents would birth up to their ideal parity ($N = n^*$), and the proportion of sons at all relative orders $j^* \leq 0$ would be equal to the natural probability of birthing a male child.*

2. *Differential stopping behavior does not alter the proportion of sons born at relative orders (j^*).*

3. *If sex-selection is resorted to at any relative order j^*, the proportion of sons born at j^* increases.*

For some intuition towards the first statement in the proposition, look at Figure 1. It considers the case of parents who vary in their ideal parity n^*, but do not engage in any of two son-targeting methods. Figure 1 Panel (A) plots the proportion of sons that would be born at each relative birth order j^*. In the absence of son-targeting, it is natural to expect that each parent would give birth to exactly n^* children and then stop, i.e. $N = n^*$. Thus, no births would not be subject to parental manipulations, and the proportion of sons at all birth orders would equal the natural probability of giving birth to a boy. Consequently, even at all relative birth orders—which is equivalent to resetting the birth order counter in this case—the proportion of sons would equal the natural probability. As a comparison exercise, to illustrate the effect of disaggregating data, Panel (B) also looks at the proportion of sons but based on how far each birth is from the total parity i.e. $k - N$. Here also, at all birth orders re-indexed relative to actual parity, the proportion of sons would equal the natural probability of giving birth to a boy.

FIGURE 1 :
Sex-ratios in the absence of son-targeting

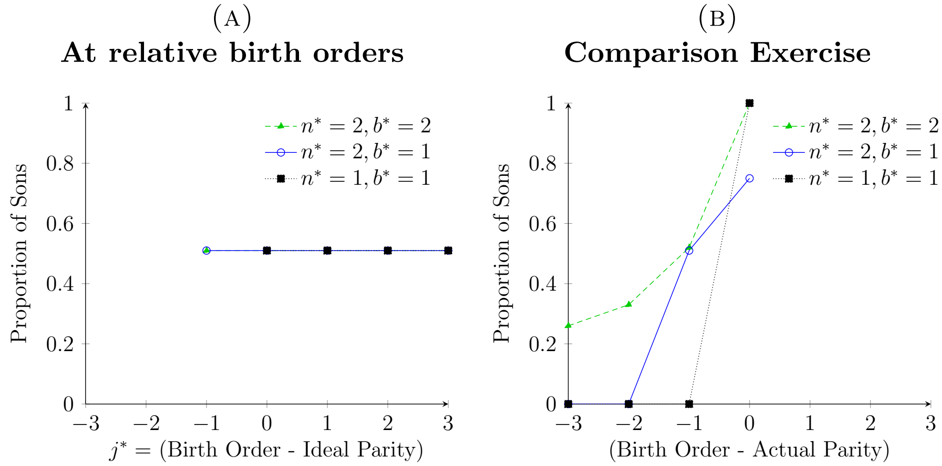


Next, consider the case of parents who only practice differential stopping. More extremely, assume these parents continue childbearing until b^* boys are secured, no matter the number of births it takes. Parents who couldn't naturally birth b^* sons in their first n^* attempts, would necessarily birth $N > n^*$ children. For them, their youngest child would always be a boy—their b^{*th} son.

Suppose $n^* = 1, b^* = 1$. See Figure 2 Panel (B). Parents who birth more than $n^* = 1$ children would only be those couldn't birth a boy at their first attempt. With $b^* = 1$, all children except the youngest child must necessarily be girls. For $n^* = 2, b^* = 1$, however, parents with a male firstborn could have their youngest child born at k = 2 be a daughter. But those have 3 children must necessarily have given birth to a daughter first. In general, the proportion of sons at lower birth orders would decrease as the family size increases, since we would be selecting on a pool of parents who decided to have more children precisely because they did not birth enough sons. It is this selection of parents into higher parity caused due to differential stopping, that can skew sex ratios at lower and higher orders *when disaggregated by actual parity.*

However, since sex-selection is not resorted to, the proportion of sons at all relative orders j^* would still be equal to the natural probability of giving birth to a boy. This is because sex is determined naturally and independently at each birth order no matter the value of n^* or k; see Figure 2 Panel (A).

FIGURE 2 :
Sex-ratios with only differential stopping



Moving onto a more realistic scenario, suppose parents cared about both the number of children and the composition. Depending on the relative importance attributed to the two, one can think of several heuristics by which parents would be willing to tolerate some deviation from n^* in order to not deviate too much from b^*, or vice-versa. For example, one heuristic is where parents unable to birth b^* boys among their first n^* children, may be willing to extend parity to $n^* + 1$ to birth a boy by sex-selecting. Many heuristics of a general form could be used: birth $n^* - m_1$ children naturally, rely on sex-selection m_2 times to birth $n^* - m_1 + m_2$ children.

With heuristics of this form, N would always be $\leq n^* - m_1 + m_2$, and the proportion of sons at relative orders $j^* = -m_1 + m_2$ would be equal to 1. Figure 3 illustrates this for when parents sex-select when at $n^* + 1$ i.e. with $m_1 = 0, m_2 = 1$; and for when parents sex-select at n^* i.e. with $m_1 = -1, m_2 = 1$.

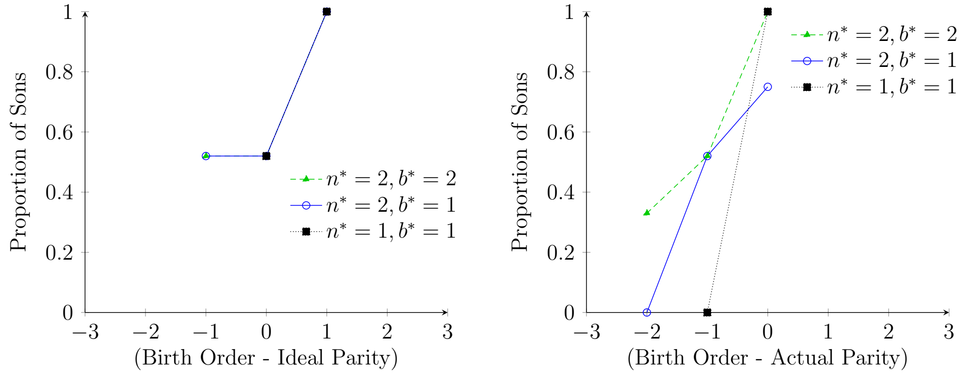
To summarize, in the absence of sex-selection, the proportion of sons at any relative order j^* will always equal the natural probability of conceiving a boy. However, only if sex-selection is resorted to at any relative order j^*, would we observe a spike in proportion of sons at j^*.

In Section IV, I define relative birth orders for 2,231,621 births across the three rounds, and compare the proportion of sons born at each relative births order to infer sex-selection. The relative orders at which parents decide to sex-select describes the heuristics of the general heuristic we are after.

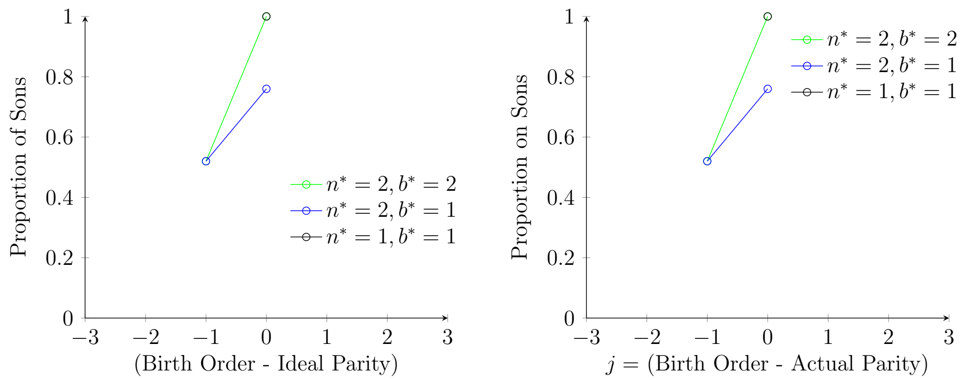
I then empirically validate the heuristic using an empirical specification that exploits the natural sex-determination system in humans, the details of which are described next.

FIGURE 3 :
Sex-ratios with sex-selection

(A)
$m_1 = 0, m_2 = 1$**: Sex-select once at** $n^* + 1$ **if no. of boys** $< b^*$



(B)
$m_1 = -1, m_2 = 1$**: Sex-select once at** n^* **if no. of boys** $< b^*$



III.C. Empirical test to validate heuristic

To validate the inferred heuristic, I test for evidence of sex-selection at each birth order and confirm if it is indeed driven by parents who the heuristic suggests would sex-select at the birth order. For the inference of sex-selection, I employ another observed statistic: the length of the inter-birth interval.

Inter-birth intervals are typically effected by parents' intended birth-spacing and the mother's fecundity. The former depends on parents' opportunity costs of time or resources, while the latter depends on maternal age, post-partum amenorrhea and contraceptive-use. Sex-selection uniquely affects the length of the observed birth interval *preceding* the birth at which sex-selection was resorted to. That is, sex-selective abortion increases the observed inter-birth interval between the last birth and the successfully manipulated male birth which follows it. This increase is on account of the inability to determine sex before 10-12 weeks of pregnancy, the time it takes for post-abortion recovery and the time till next conception (Pörtner, 2022). The more abortions done by a parent to eliminate female births, the longer would be the inter-birth interval.

Thus, sex-selection that results in a live male birth recorded at any birth order k would also render the birth interval *preceding* the k^{th} birth to be longer.[14] Considering the temporal ordering of birth interval and the natural (and legal) revelation of the sex of the child at birth (at the end of this interval); for any correlation between the length of a birth interval and the sex of the child born at the end of it, the direction of effect is obvious: the sex recorded at the end of the interval is effected by the "practices" of parents during that interval. Studying how the proportion of male births varies with the length of the preceding birth interval therefore provides insights into the dynamics of fertility choices made between two live births.

Figure 4 plots the proportion of male births at the second birth order, conditional on the length of the preceding birth interval. The figure has separate subplots for births following different histories. If sex at consecutive births were determined independently in the absence of sex-selection, we would expect the plots to be identical for all histories and be horizontal at the value of the natural probability of male births. Even if the relation between probability of birth a boy and the length of the preceding interval were non-linear, as long as sex of the consecutive child is determined independently, we would expect the plots to be, at the very least, identical for different histories. Instead, Figure 4 suggests that mothers with firstborn girls are disproportionately more likely to give birth to sons at the second order and more so after longer birth intervals. This trend is consistent across the three NFHS rounds we examine. Interesting, the increased tendency to birth sons also reverses after the inter-birth interval exceeds 5-7 years. This suggests that some parents at birth history `F` engage in sex-selection, with many of these parents taking-back their decision to sex-select as the inter-birth interval gets "too long".

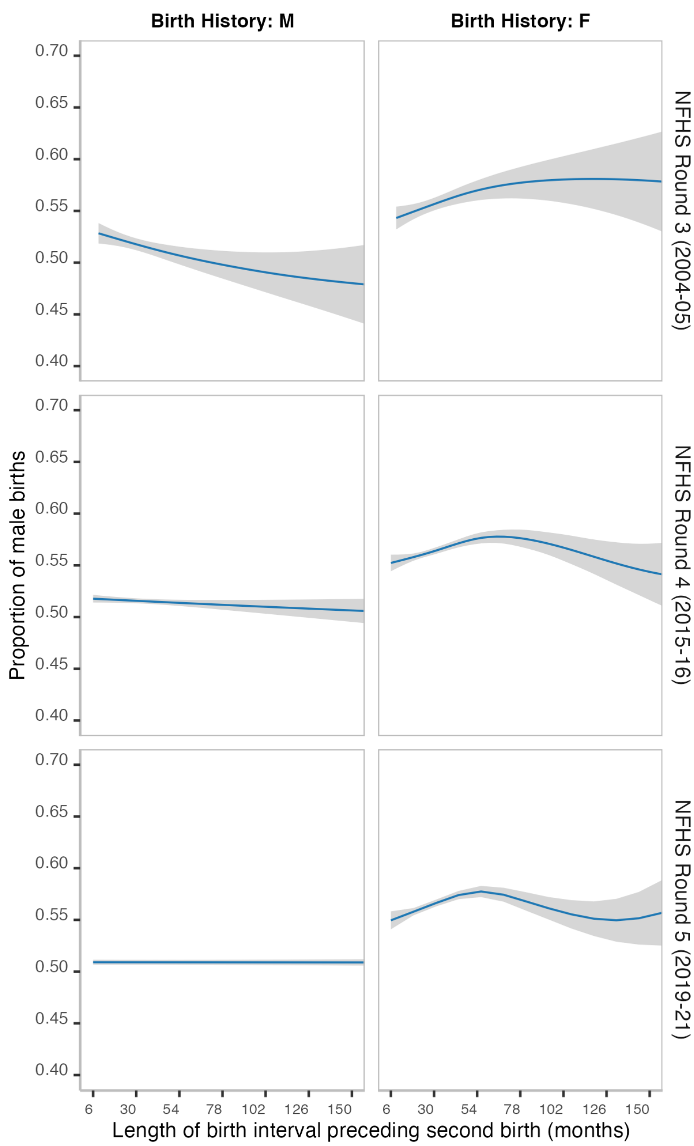
Likewise, Figures 5 and 6 plot the proportion of male births against the length of the preceding interval for children born at the third and fourth birth orders respectively. Once again, following histories `FF` and `FFF` the proportion of male births increases steadily with the length of the preceding interval.

Appendix Figures A1 - A3 repeat these comparisons at the second birth order for different levels of mothers' education, different places of residence, and different geographical regions. Appendix Figures A4 - A6 do the same comparisons for births at the third birth order. Our observations of boys being born after longer intervals following birth histories with no or few sons, hold consistently across both urban and rural areas, and for all levels of maternal education.

A comparison by geographical region — also synonymous with cultural divisions in India — echoes the well documented regional disparity in sex-ratios and gender relations in India

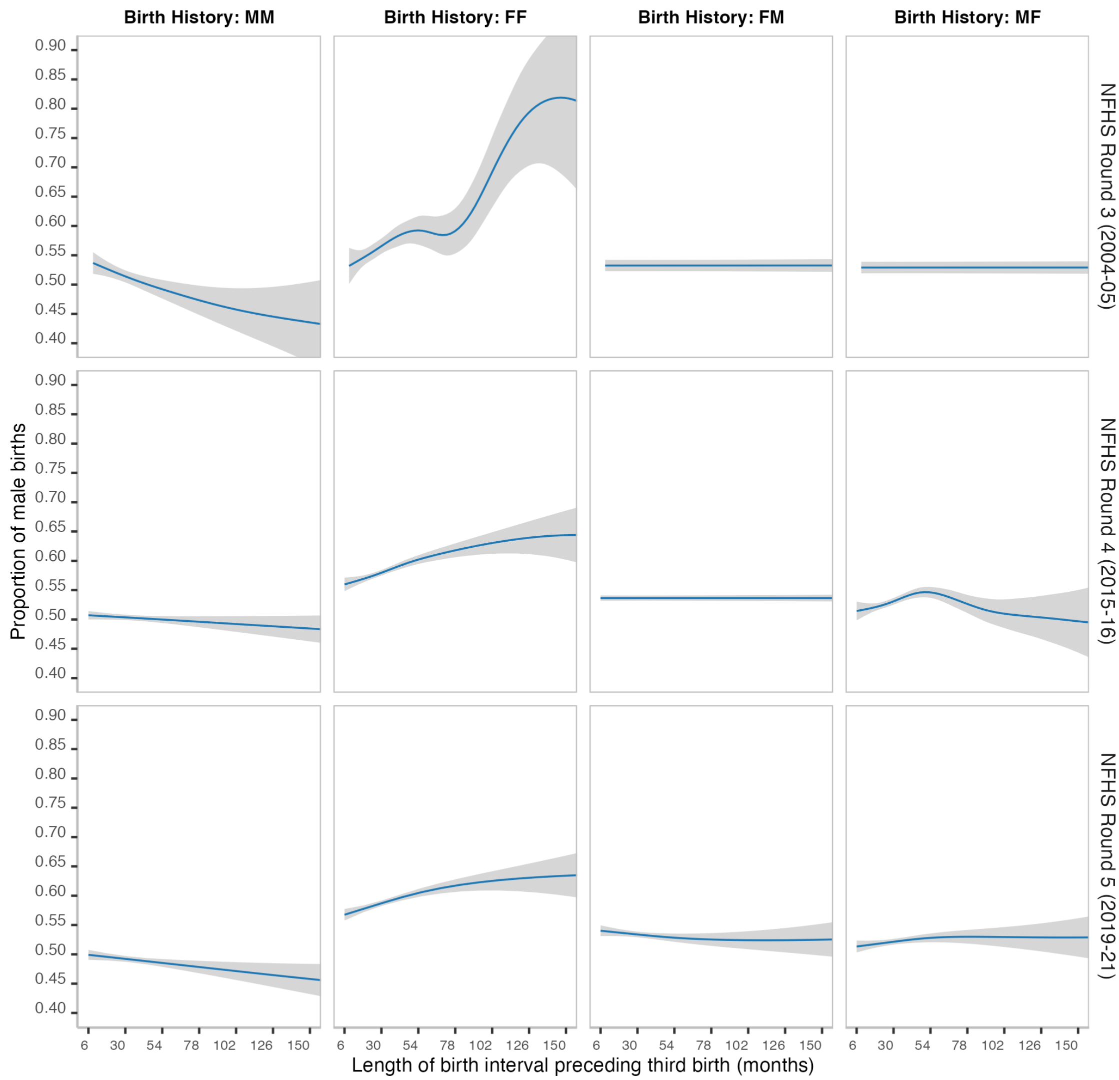
[14]This is true if there is no reversal in the decision to sex-select (Pörtner, 2022, 2016). In the case of reversals in the decision to sex-select, we would also observe girls being born after failed sex-selective attempts and hence after long intervals. The net effect on the direction of correlation between sex of a newborn and the preceding interval would then depend on the specific composition of the 'types' of sex-selectors: those who would reverse their decision after failed attempts, and those who do not. Similarly, if sex-selecting parents anticipate a long inter-birth interval and begin their attempts at conception sooner in anticipation, then also the net effect on the direction of correlation would be ambiguous. Therefore, correlation between sex and preceding birth interval confirms sex-selection, but an absence in correlation doesn't deny its use.

FIGURE 4 :
Variation in probability of male births (at the second birth order) at varying lengths of the preceding birth interval



Notes: This figure plots kernel-smoothed proportion of male births at the second birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories.

FIGURE 5 :
Variation in probability of male births (at the third birth order) at varying lengths of the preceding birth interval



Notes: This figure plots kernel-smoothed proportion of male births at the third birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories.

(Dyson & Moore, 1983). North-western states that include Rajasthan, Punjab and Haryana, along with states in Western and Central India, such as Gujarat, exhibit strong sex-selecting behavior following histories `F` and `FF`; as is also reflected in the low female-to-male ratios in these regions (see Appendix Figures A3 and A6). In contrast are the Southern and North-Eastern states — typically with more balanced sex-ratios and higher agency for women — that display no such trend in comparison.

Exploiting this effect of sex-selection on the preceding birth intervals, I employ a difference-in-difference specification to infer sex-selection at each birth order.

At each birth order k, I compare correlations between sex of the child born to mother m and the interval preceding its birth (denoted by '$preI_{mk}$'), for various birth histories $h \in H_k$ referenced against the history with all male births. Here, H_k denotes the set of all sex-sequences until birth order k. For example, for births at the third birth order: $k = 3$ and $H_k = \{\texttt{MM, FF, MF, FM}\}$.

The following specification is used:

$$\begin{aligned} preI_{mk} = const + \sum_{h \in H_k \setminus h_r} \alpha_h \{\text{hist}_{mk}\text{:h}\} + \beta \{\text{sex}_{mk}\text{:M}\} + \\ \sum_{h \in H_k \setminus h_r} \gamma_h \{\text{hist}_{mk}\text{:h} \times \text{sex}_{mk}\text{:M}\} + \theta' X_{mk} + u_{mk} \end{aligned} \tag{4}$$

$$\begin{aligned} h_r &: \text{birth history with } k-1 \text{ male children} \\ \text{hist}_{mk}\text{:h} &: 1 \text{ if birth history until order } k-1 \text{ for mother } m \text{ is } h, 0 \text{ otherwise} \\ \text{sex}_{mk}\text{:M} &: 1 \text{ if } k^{th} \text{ child of mother } m \text{ is male, } 0 \text{ otherwise} \\ X_{mk} &: \text{vector of covariates} \end{aligned}$$

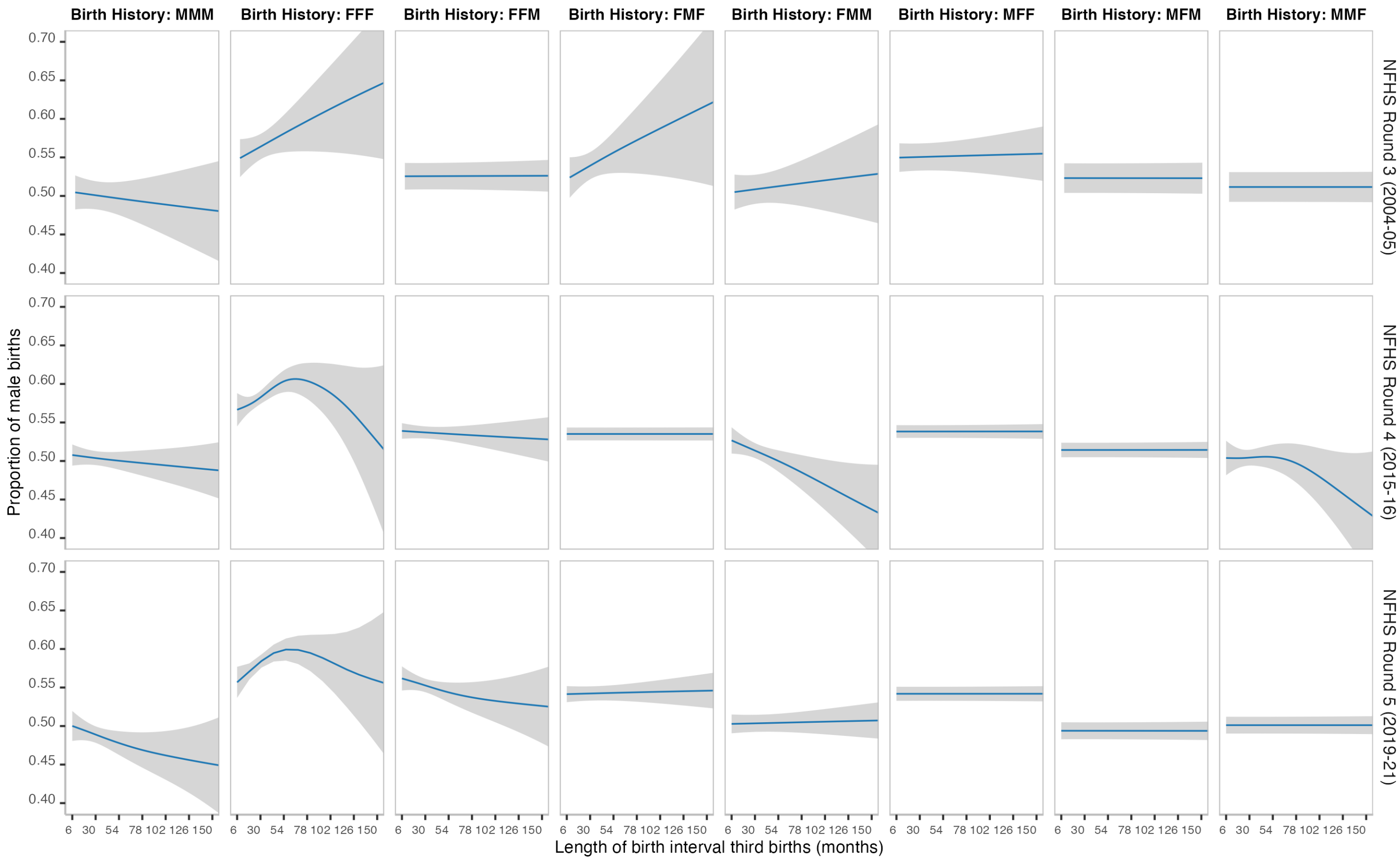
It estimates the difference between intervals preceding male births to those preceding female births for a given history h, and then differences this difference relative to that for a reference history h_r with all male births. Here, the coefficients of interest are γ_h for all histories $h \in H_k \setminus h_r$. Positive and significant coefficients indicate sex-selection. In the case of reversals in the decision to sex-select, we would also observe girls being born after failed sex-selective attempts and hence after long intervals. The coefficients in such a case would be negative and significant. Nevertheless, a significant relation between sex and the length of the preceding birth interval following only select histories confirms sex-selection.

The identifying assumption here is that *even if* there are any unaccounted unobservables (such as geography, ethnicity, socio-economic or environmental factors) that could skew the natural probability of birthing a son *and* influence the length of the preceding interval; they should confound the relationship between sex and preceding interval, if at all, for *any* birth history. Since the incentives to sex-select vary with birth history but the effect of confounding variables do not, any significant difference-in-difference in length of the preceding interval by the sex of the newborn, must be attributed to parental manipulation.

I then use this difference-in-difference specification to the empirically validate the identified heuristic, as follows. For each birth order k, I partition the set of mothers into whose ideal parity preferences are such that the heuristic suggests they would sex-select at k, and into all others. Whenever sex-selection is resorted to any birth order, the heuristic is validated if sex-selection at the order—as indicated by the difference-in-difference—is driven by mothers in the first partition.

Figure 6 :

Variation in probability of male births (at the fourth birth order) at varying lengths of the preceding birth interval



Notes: This figure plots kernel-smoothed proportion of male births at the fourth birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories.

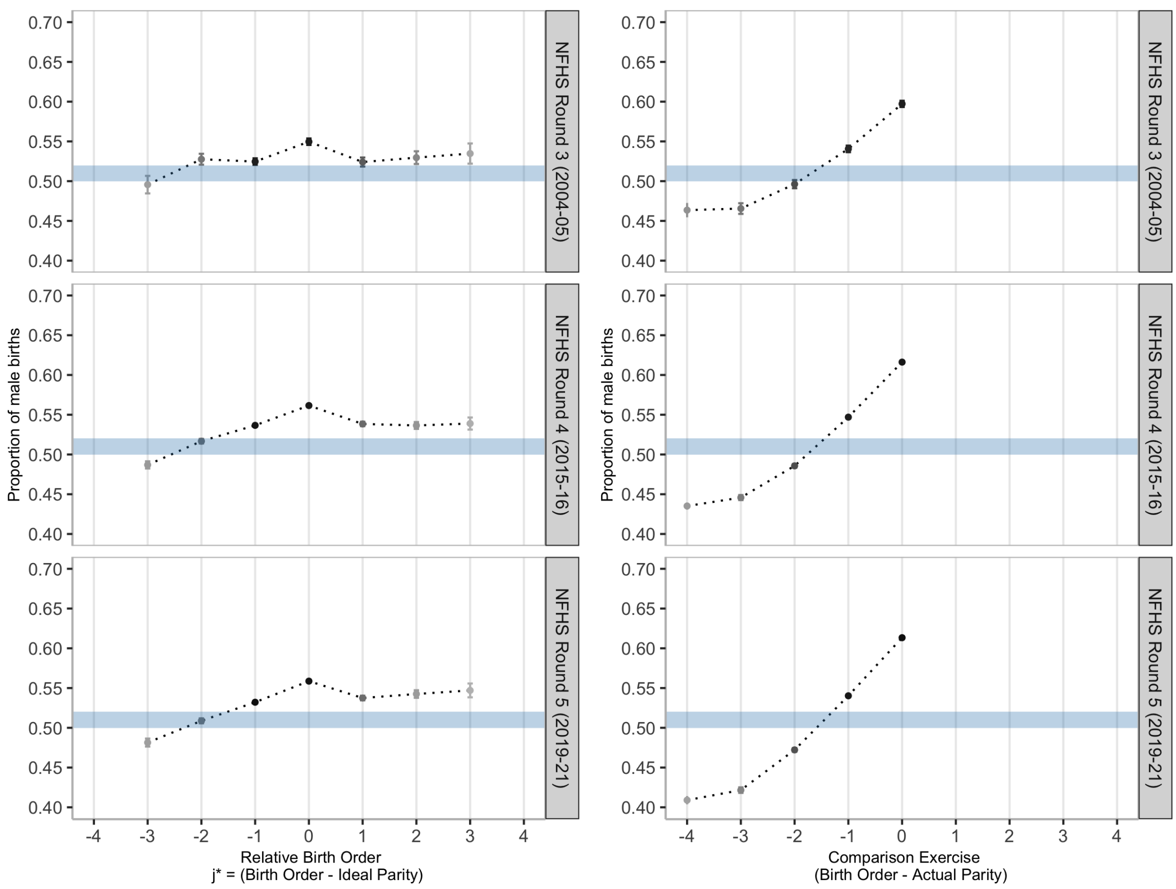
IV. EMPIRICAL FINDINGS

IV.A. Heuristic of Sex-Selection

With relative birth orders defined for all births in the sample, let us first examine the proportion of sons at each relative birth order. Here, analysis is restricted to 2,231,621 births across the three rounds, born to mothers who report a strictly positive ideal parity.

The left panel in Figure 7 plots the proportion of sons born at various relative birth orders, disaggregated by the sample round. The right panel of the figure plots the same, but at birth orders defined relative to the actual parity, as a comparison exercise. The dissimilarity between the two panels offers another indication that respondents do not simply claim their ideal parity to be the same as their actual parity.

Figure 7 :
Proportion of male births by relative birth order



Notes: This left panel of this figure plots the proportion of male births at various relative birth orders. The right panel—as a comparison exercise—plots proportion of male births at birth orders re-indexed by the mother's actual parity. The shaded horizontal bar indicates the range 0.50-0.52 denoting the natural probability of male births Orzack et al. (2015). The color gradient of the markers indicate the share of births in the sample contributing to the statistic at the relative birth order, with darker shades representing higher shares.

Recall that sex-selection is indicated by high proportions of sons at relative birth orders. The left panel of Figure 7 suggests that most parents who sex-select, do so at the relative birth order $j^* = 0$, that is, when they are at their ideal parity. Births that exceed the ideal parity, i.e. those at relative birth order $j^* > 0$, are typically not subject to sex-selective abortions. The right panel of the figure confirms that indeed the last birth is predominantly male, with a selection of parents (those with fewer sons) deciding to have more children and stopping after a male birth.

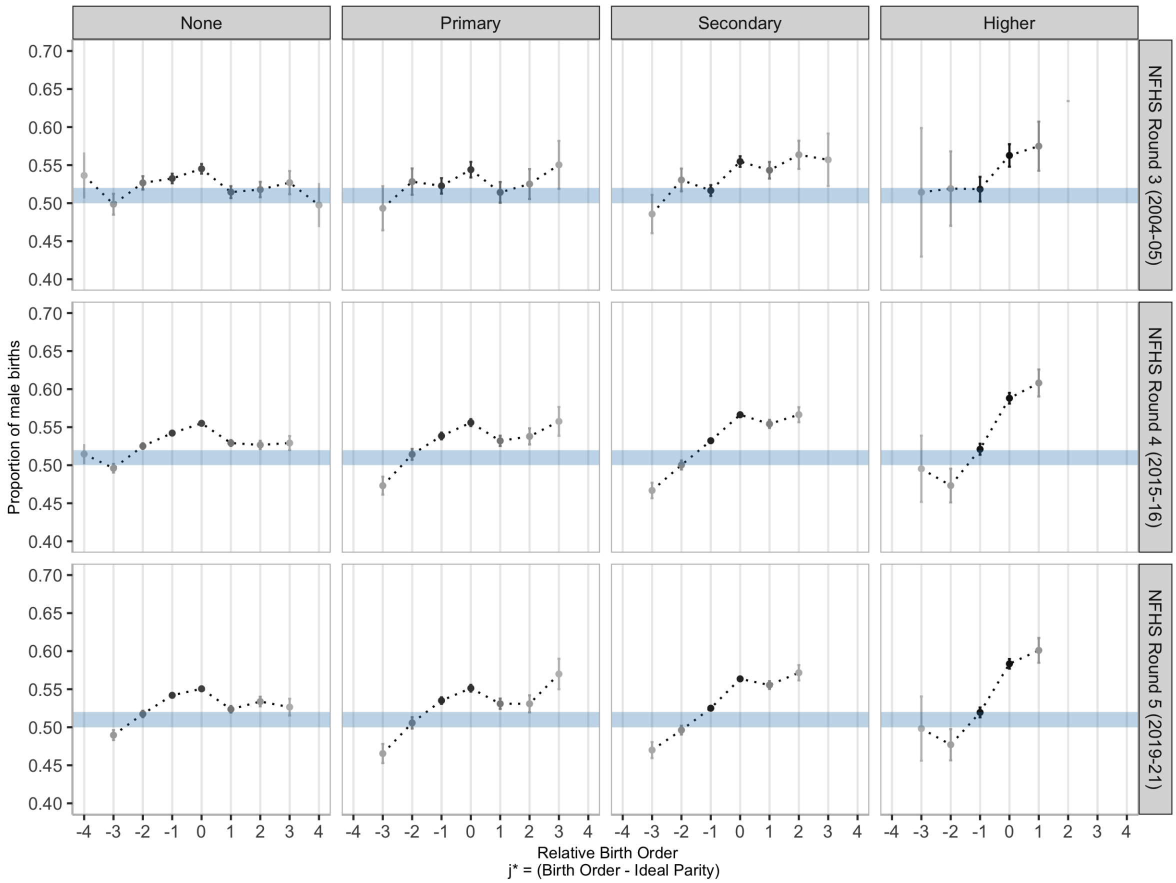
Appendix Figure A7 plots the proportion of sons born at various relative birth orders, disaggregated jointly by the geographical region and the sample round. Consistent with Appendix Figures A3 and A6, evidence of sex-selection is clear and pronounced for states in the North-West and in the West & Central parts of India as indicated by a notable increase in the proportion of sons at the relative order $j^* = 0$. In contrast, states in South or East India show no such empirical pattern.

To gather more insights on who decides to sex-select, Figure 8 examines how the empirical pattern varies with mother's level of education. Figure 8 suggests that resorting to sex-selection at the relative birth order $j^* = 0$ (i.e. at the ideal parity) is common to mothers at all levels of education. Two important aspects are worth noting here. First, mothers with higher-level education rarely exceed their ideal parity. Most mothers who exceed their parity are primary-educated or uneducated. The costs associated with positive deviations from ideal parity, therefore, likely increase with the level of education. But the heuristics of *when* to sex-select, are largely identical.

Second, while most higher-level and secondary-level-educated mothers follow this general heuristic, some also decide to exceed their ideal parity by one and sex-select. Educated parents who exceed their ideal parity are much more likely to sex-select, compared to uneducated mothers, who if they exceed their ideal parity, are mostly following differential stopping. This increased tendency of educated mothers to sex-select, also previously noted in (Bhalotra & Cochrane, 2010; Pörtner, 2016, 2022), comes in stark contrast with weakened son-preference noted among the educated (Jayachandran, 2017). There are two possible reasons for this. First, higher education may be correlated with increased access or means to sex-select. Second, if maternal education is correlated with lower ideal parity, it could then intensify any tendency to sex-select. This is because achieving a desired composition with a small parity is probabilistically harder based on the natural odds. Which of these two reasons is driving the relation between maternal education and sex-selection is an important area for further research. A strikingly similar empirical pattern is generated as I vary the wealth quintile; see Appendix Figure A8.

Nevertheless, the increased tendency to sex-select when at the ideal parity is consistent across the three rounds and at all levels of education. I therefore hypothesize a general heuristic whereby a parent with an ideal parity of n^* and an unmet desire for sons would sex-select, if at all, at the birth order $k = n^*$ (i.e. at $j^* = 0$). There may be exceptions to this heuristic, but it defines the strategy of a large share of Indian parents.

FIGURE 8 :
Proportion of male births at relative birth orders, by mother's education



Notes: This figure plots the proportion of male births at various relative birth orders, by mother's education. The vertical columns represent the level of mother's education and the horizontal panels specify the NFHS round. The shaded horizontal bar indicates the range 0.50-0.52 denoting the natural probability of male births Orzack et al. (2015). The color gradient of the markers indicate the share of births in the sample contributing to the statistic at the relative birth order, with darker shades representing higher shares.

IV.B. Empirical Validation of the heuristic

To empirically validate the heuristic I use the specification in Equation (4) to test if the sex-selection at any birth order k is driven by parents with a preferred ideal parity of $n^* = k$. The controls in the specification include mother's age at previous birth, an indicator for rural place of residence, wealth quintile, NFHS round-specific fixed effects and state fixed effects.

Table 3 reports the estimates for births at the second birth order. Table 3 column (1) confirms the use of sex-selection at the second birth order, where boys are born after longer intervals compared to girls following the history `F`. A negative coefficient on hist : `F` reaffirms the finding of parents beginning conception for the successive birth sooner after the birth of a female firstborn (Jayachandran & Kuziemko, 2011; Javed & Mughal, 2020).

Columns (2) and (3) in Table 3 re-estimate the specification in two different sub-samples. Column (2) reports the estimates for mothers who the heuristic suggests are most likely to sex-select at the second birth order, i.e. mothers with $n^* = 2$. Column (3) repeats the exercise for a placebo sample of mothers with $n^* > 2$ who we do not expect would resort to sex-selection at the second order. As expected, we find the coefficient of the interaction term hist: `F` $\times$ sex: `M` to be significant in the sub-sample suggested by the heuristic, and statistically insignificant for the placebo sub-sample.

TABLE 3 :
Empirical validation of the heuristic at the second birth order

	Dependent variable: Interval preceding **second** birth		
	All mothers with $N \geq 2$	Mothers with $n^* = 2$ (Heuristic)	Mothers with $n^* > 2$ (Placebo)
	(1)	(2)	(3)
hist: F	-1.466^{***}	-1.297^{***}	-1.197^{***}
	(0.071)	(0.099)	(0.104)
sex: M	-0.139^{**}	-0.309^{***}	0.065
	(0.068)	(0.090)	(0.106)
hist: F $\times$ sex: M	0.550^{***}	0.610^{***}	-0.103
	(0.098)	(0.132)	(0.147)
Constant	29.809^{***}	28.833^{***}	34.815^{***}
	(0.245)	(0.316)	(0.462)
Observations	732,947	435,254	254,246
Note:			*p<0.1; **p<0.05; ***p<0.01

This table reports the difference-in-difference estimates based on the specification in Equation (4) at the second birth order. Specification (1) reports the estimates for all mothers who give birth to at least two children. Specifications (2) and (3) report the estimates for mothers who according to our heuristic are expected to sex-select i.e. those whose ideal parity is equal to the given birth order ($n^* = 2$) and a placebo sample. The controls include mother's age at previous birth, rural, wealth quintile, NFHS round-specific fixed effects and state fixed effects. Robust standard errors included in parentheses.

Table 4 does the same exercise at the third birth order to validate the heuristic. Once again, I find evidence of sex-selection only in the sub-sample suggested by the heuristic, and after the birth history `FF`. Appendix Table B2 reports the estimates for the fourth birth order. We find evidence of sex-selection following histories `FFF`, `FMF`, and `MFF`. But at the fourth order, sex-selection is not driven by mothers reporting an ideal parity of four. These are likely parents who exceed their ideal parity and are sex-selecting. It is important to exercise caution when interpreting results for specific histories at high birth orders because the sample size driving these results falls dramatically. Moreover, as we analyze higher birth orders ($k \geq 4$), our sub-sample selects on parents who are willing to have at least k children. Thus, as k increases, we would be selecting on parents who either have a large ideal parity, or for whom exceeding their ideal parity is probably not much of a concern.

TABLE 4 :
Empirical validation of the heuristic at the third birth order

	Dependent variable: Interval preceding **third** birth		
	All mothers with $N \geq 3$	Mothers with $n^* = 3$ (Heuristic)	Mothers with $n^* > 3$ (Placebo)
	(1)	(2)	(3)
hist: FF	-1.347^{***}	-1.651^{***}	-1.758^{***}
	(0.133)	(0.230)	(0.265)
hist: FM	-0.746^{***}	-0.463^{**}	-0.720^{***}
	(0.132)	(0.219)	(0.268)
hist: MF	-1.238^{***}	-1.022^{***}	-1.149^{***}
	(0.131)	(0.217)	(0.267)
sex: M	-0.463^{***}	-0.424^{**}	-0.129
	(0.132)	(0.214)	(0.278)
hist: FF × sex: M	1.641^{***}	1.522^{***}	0.504
	(0.179)	(0.311)	(0.354)
hist: FM × sex: M	0.392^{**}	-0.026	-0.079
	(0.180)	(0.298)	(0.373)
hist: MF × sex: M	0.749^{***}	0.352	0.596
	(0.180)	(0.296)	(0.369)
Constant	28.206^{***}	27.713^{***}	30.806^{***}
	(0.339)	(0.635)	(0.870)
Observations	425,344	139,090	91,972
Note:			*p<0.1; **p<0.05; ***p<0.01

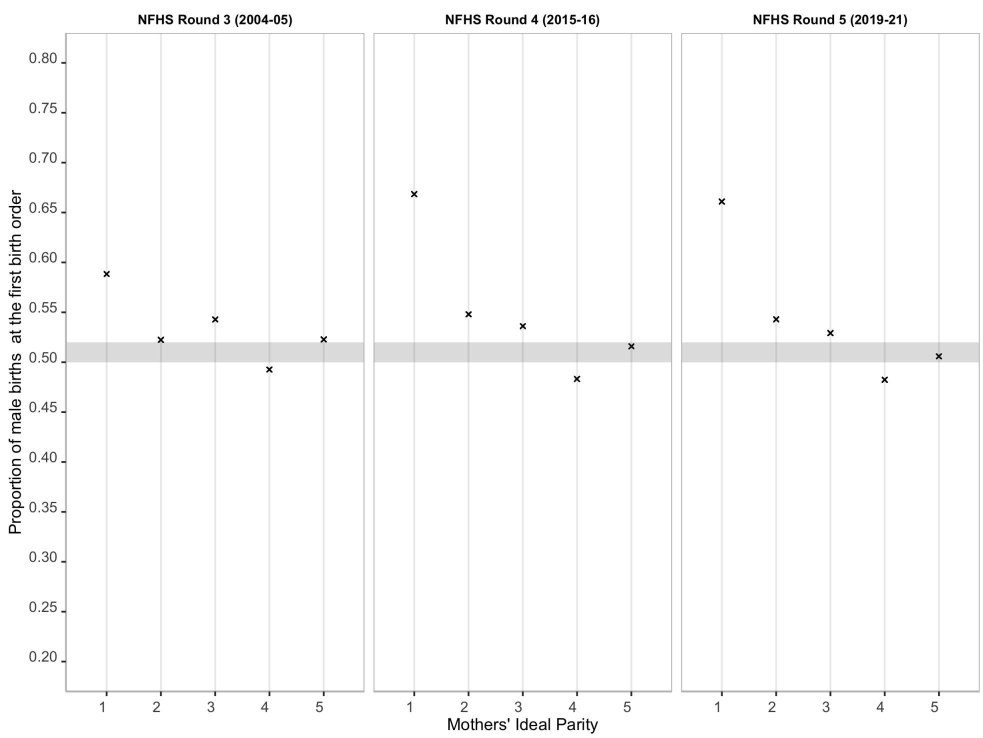
This table reports the difference-in-difference estimates based on the specification in Equation (4) at the third birth order. Specification (1) reports the estimates for all mothers who give birth to at least three children. Specifications (2) and (3) report the estimates for mothers who according to our heuristic are expected to sex-select i.e. those whose ideal parity is equal to the given birth order ($n^* = 3$) and a placebo sample. The controls include mother's age at previous birth, rural, wealth quintile, NFHS round-specific fixed effects and state fixed effects. Robust standard errors included in parentheses.

IV.C. Sex-selection at the first birth order

The discussion so far has only been on sex-selection at birth orders strictly greater than one. This is because there are no histories before the first birth to exploit for the difference-in-difference identification. Moreover, the interval preceding the first birth, that is, the time since marital union to first birth is conceptually very different from inter-birth intervals.

Previous research that relies on the sex of the firstborn being random defends the assumption based on sex-ratios at the first order being within a 'normal' range. Disaggregation of the statistic by mothers' ideal parity shows that mothers with an ideal parity of one are disproportionately more likely to birth a male child at the first birth order (Figure 9). Further, for mothers with an ideal parity of one, the proportion of male births at the firth birth order also increases with the length of the interval between marital union and first birth; see Figure 10. This is not observed for mothers who report higher ideal parity.

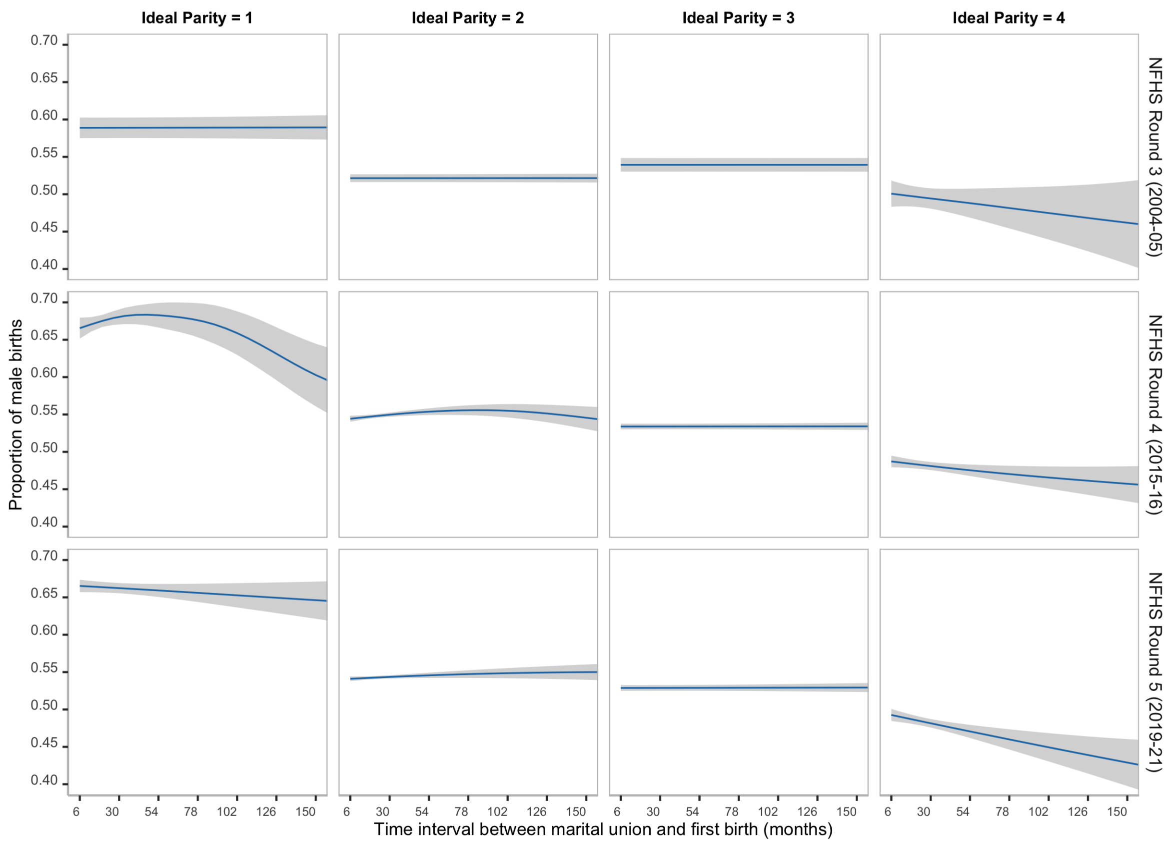
FIGURE 9 :
Proportion of male births at the first birth order



Notes: This figure plots the proportion of male births at the first birth order, disaggregated by mothers' ideal parity and the NFHS round. The shaded horizontal bar indicates the range 0.50-0.52 denoting the natural probability of male births Orzack et al. (2015).

To empirically confirm sex-selection at the first birth order, I use the heuristic. Specifically, according to the heuristic, parents most likely to sex-select at the first birth order are those with ideal parity of one or more generally, those with $n^* = 1$. I employ a difference in difference specification that compare the length of the interval since marital union to the first birth, by the sex of the firstborn and between sub-samples that the heuristic predicts

FIGURE 10 :
Variation in probability of male births (at the first birth order) at varying lengths of the preceding birth interval



Notes: This figure plots kernel-smoothed proportion of male births at the first birth order, conditional on the length of preceding interval between marital union and first birth. Each panel plots the conditional proportions for mothers with different values of ideal parity.

will sex-select against one that won't. Consider the following specification:

$$\begin{aligned} log(I_{m1}) = const + \alpha\mathbb{1}\{\text{sub-sample qualifier}\} + \beta\{\text{sex:M}\} + \\ \gamma\mathbb{1}\{\text{sub-sample qualifier}\} \times \{\text{sex:M}\} + \theta' X_{m1} + u_{m1} \end{aligned} \quad (5)$$

$$\begin{aligned} m &: \text{index for mother} \\ I_m &: \text{interval between marriage/union and the first birth for } m \\ \text{sex:M} &= 1 \text{ if first child is male, 0 otherwise} \\ \text{sub-sample qualifier} &: \mathbb{1}\{n^* = 1\} \text{ for the heuristic, and } \mathbb{1}\{n^* > 1\} \text{ as placebo} \\ X_m &: \text{vector of covariates} \end{aligned}$$

Consider two sub-samples: one of mothers that the heuristic suggests would sex-select at the first birth order, i.e. those with an ideal parity of one ($n^* = 1$); and another placebo sub-sample of mothers with ideal parity greater than one. Table 5 reports the estimates for the specification.

Specification (1) reports the estimates for all mothers in our sample. Specifications (2) and (3) report the estimates using the sub-sample qualifiers for the heuristic and placebo.

Table 5 :
Sex-selection at the first birth order, using the heuristic

	Dependent variable:		
	Interval preceding **first** birth (since marriage/union)		
	Baseline	Heuristic	Placebo
	(1)	(2)	(3)
sex: M	-0.103^{*}	-0.201^{***}	-0.303
	(0.053)	(0.054)	(0.230)
Heuristic: $\mathbb{1}\{n^{*}=1\}$		6.698^{***}	
		(0.258)	
Placebo: $\mathbb{1}\{n^{*}>1\}$			-4.787^{***}
			(0.188)
sex M $\times$ Heuristic: $\mathbb{1}\{n^{*}=1\}$		-1.073^{***}	
		(0.309)	
sex M $\times$ Placebo: $\mathbb{1}\{n^{*}>1\}$			0.014
			(0.236)
Constant	56.126^{***}	56.629^{***}	61.203^{***}
	(0.289)	(0.291)	(0.352)
Observations	742,362	737,404	737,404
Note:			*p<0.1; **p<0.05; ***p<0.01

This table reports the difference-in-difference estimates based on the specification in Equation (5) at the first birth order. Specification (1) reports the estimates for all mothers in our sample. Specifications (2) and (3) report the estimates using the sub-sample qualifiers for the heuristic and placebo, respectively. The controls include mother's age at previous birth, rural, wealth quintile, NFHS round-specific fixed effects and state fixed effects. Robust standard errors included in parentheses.

Table 5 column (2) reports a significant correlation between the sex of the first child and the length of the interval preceding the first birth, for mothers with an ideal parity of one (relative to others). The negative coefficient suggests that there is likely some reversal in the decision to sex-select after failed attempts. No such correlation is observed for the placebo group of mothers with an ideal parity greater than one.

V. CONCLUSION

The prevalence of son-preference, as well as the practice of differential stopping and sex-selection, are well documented. This paper contributes in characterizing parental son-targeting strategies.

I characterize Indian parents' son-targeting strategy in the form of a general heuristic that describes *when* parents decide to sex-select. To identify the heuristic, I define the relative birth order for each birth, indicating how far the birth is from the mother's ideal parity. I then examine the probability of male birth at these various relative birth orders. Doing so, allows a comparison of parental strategies across families of different size, and also resolves the issue of endogeneity of male-skewed sex-ratio at the last birth order.

I find a significant increase in the proportion of male births among parents (with no or few sons) at their ideal parity, suggestive of a heuristic whereby parents resort to sex-selection at their ideal parity to avoid exceeding it. Those that don't sex-select at their ideal parity, mostly continue childbearing using differential stopping. If at higher parities they are still unable to birth a male child, some consider sex-selection. I validate the general heuristic of sex-selection at the ideal parity by exploiting the non-relation between the sex of a newborn and the length of the inter-birth interval preceding its birth. The use of sex-selection increases the observed preceding inter-birth interval rendering the preceding intervals to be longer for male than female births. To account for the confounding effects of any unobserved factors, I adopt a difference-in-difference specification that compares the correlation between sex and preceding interval for parents with different birth histories. Since the incentive to sex-select varies with birth history but the effect of omitted variables should not, any significant difference in correlation in length of the preceding interval by the sex of the newborn (relative to a reference history), must be attributed to parental manipulation.

I make three important findings. The first is the heuristic itself; suggesting that parents typically resort to sex-selection at the ideal parity. This is pertinent since there has been a declining trend in desired parity. This means that parents are likely to now sex-select at lower birth orders. It translates not only to shift in the practice of sex-selection from higher to lower birth orders, but could also increase the number of sex-selectors since achieving a desired sex-composition is probabilistically harder at lower parities. Second, identification of the heuristic also makes way in finding evidence for sex-selection at the first birth order. And lastly, we not only document sex-selection at ideal parity, but also note a tendency to reverse this decision following unsuccessful attempts.

These findings warn against any simplistic expectations about the use of sex-selective methods declining with economic development or increased maternal education. After the failure of the ban on pre-natal sex-determination under the PNDT Act, the Indian government has also explored using financial incentives to encourage parents to value girls. While this backfired in the case of the Devi Rupak program in Haryana (Anukriti et al., 2022), it worked with the Dhanlakshmi scheme (Biswas et al., 2023). Such schemes are however both costly, and likely to only nudge low income households (Jayachandran, 2024). A potential alternative, based on the heuristic, is to strengthen the enforcement of PNDT Act using random audits at the level of delivery clinics or hospitals, especially at lower birth orders.

References

Anukriti, S., Bhalotra, S., & Tam, E. H. (2022). On the Quantity and Quality of Girls: Fertility, Parental Investments, and Mortality. *The Economic Journal*, *132*.

Baland, J.-M., Cassan, G., & Woitrin, F. (2020). The Stopping Rule and Gender selective mortality: World Evidence. *CEPR Discussion Papers*.

Baland, J.-M., Cassan, G., & Woitrin, F. (2023). Sex-Selective Abortions and Instrumental Births as the two faces of the Stopping Rule. New measures and World Evidence. *CEPR Discussion Papers*.

Basu, D., & Jong, R. D. (2010). Son targeting fertility behavior: Some consequences and determinants. *Demography*, *47*, 521-536.

Bhalotra, S., & Cochrane, T. (2010). Where Have All the Young Girls Gone? Identification of Sex Selection in India. *IZA Discussion Papers, No. 5381*.

Biswas, N., Cornwell, C., & Zimmermann, L. V. (2023). The power of lakshmi: Monetary incentives for raising a girl. *Journal of Human Resources*.

Clark, S. (2000). Son Preference and Sex Composition of Children: Evidence from India. *Demography*, *37*(1), 95-108.

Das, N. (1987). Sex Preference and Fertility Behavior: A Study of Recent India Data. *Demography*, 517-530.

Dyson, T., & Moore, M. (1983). On Kinship Structure, Female Autonomy, and Demographic Behavior in India. *Population and Development Review*, 35-60. doi: 10.2307/1972894

Javed, R., & Mughal, M. (2020). Preference for Boys and Length of Birth Intervals in Pakistan. *Research in Economics*, 140-152.

Jayachandran, S. (2015). The Roots of Gender Inequality in Developing Countries. *Annual Review of Economics*, 63-88.

Jayachandran, S. (2017). Fertility Decline and Missing Women. *American Economic Journal: Applied Economics*, *9*(1), 118-139.

Jayachandran, S. (2024). Ten facts about son preference in india. *India Policy Forum, 2023*.

Jayachandran, S., & Kuziemko, I. (2011). Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India. *The Quarterly Journal of Economics*, *126*, 1485-1538.

Jayachandran, S., & Pande, R. (2017). Why Are Indian Children So Short? The Role of Birth Order and Son Preference. *American Economic Review*, 2600-2629.

Jha, P., Kumar, R., Vasa, P., Dhingra, N., Thiruchelvam, D., & Moineddin, R. (2006). Low Male-to-Female Sex Ratio of Children Born in India: National Survey of 1· 1 Million Households. *The Lancet*, *367*, 211-218.

Orzack, S. H., Stubblefield, J. W., Akmaev, V. R., Colls, P., Munné, S., Scholl, T., ... Zuckerman, J. E. (2015). The human sex ratio from conception to birth. *Proceedings of the National Academy of Sciences*, *112*, E2102-E2111.

Pande, R. P., & Astone, N. M. (2007). Explaining Son Preference in Rural India: The Independent Role of Structural versus Individual Factors. *Population Research and Policy Review*, *26*(1), 1-29.

Pörtner, C. C. (2016). Sex-Selective Abortions, Fertility and Birth Spacing. *World Bank Policy Research Working Paper, no. WPS 7189.*.

Pörtner, C. C. (2022). Birth Spacing and Fertility in the Presence of Son Preference and Sex-Selective Abortions: India's Experience Over Four Decades. *Demography*, *59*(1), 61-88.

Rasul, I. (2008). Household Bargaining Over Fertility: Theory and Evidence From Malaysia. *Journal of Development Economics*, 215-41.

Sen, A. (1990). More Than 100 Million Women Are Missing. *The New York Review of Books*.

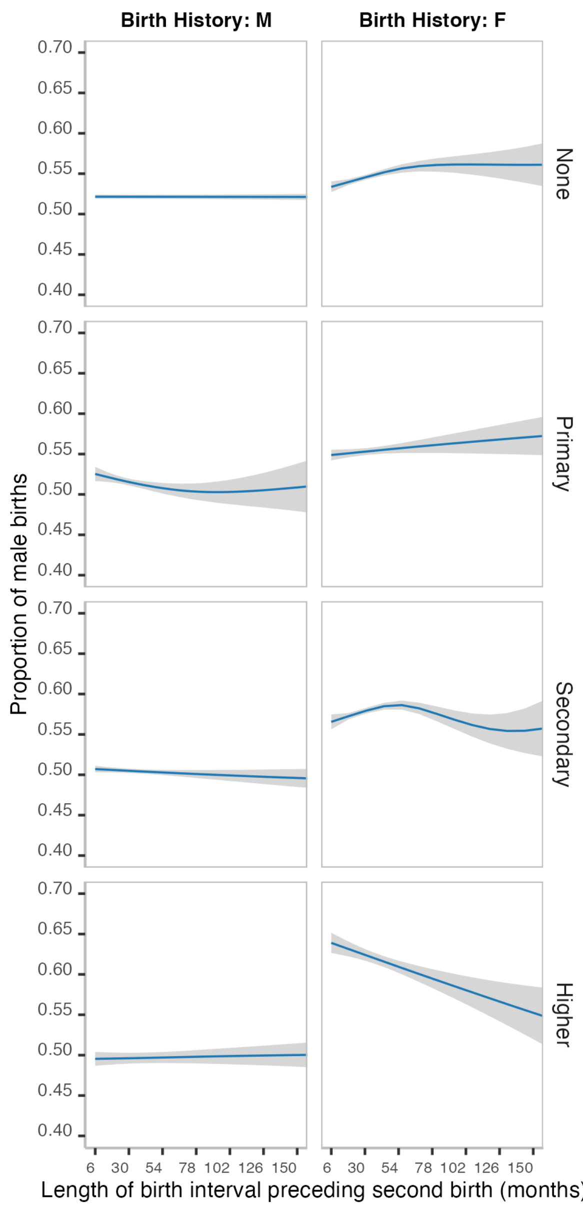
Sen, A. (1992). Missing Women. *British Medical Journal*, *304*, 587–588.

Tables created in R using stargazer v.5.2.2 by Marek Hlavac, Harvard University.

APPENDIX

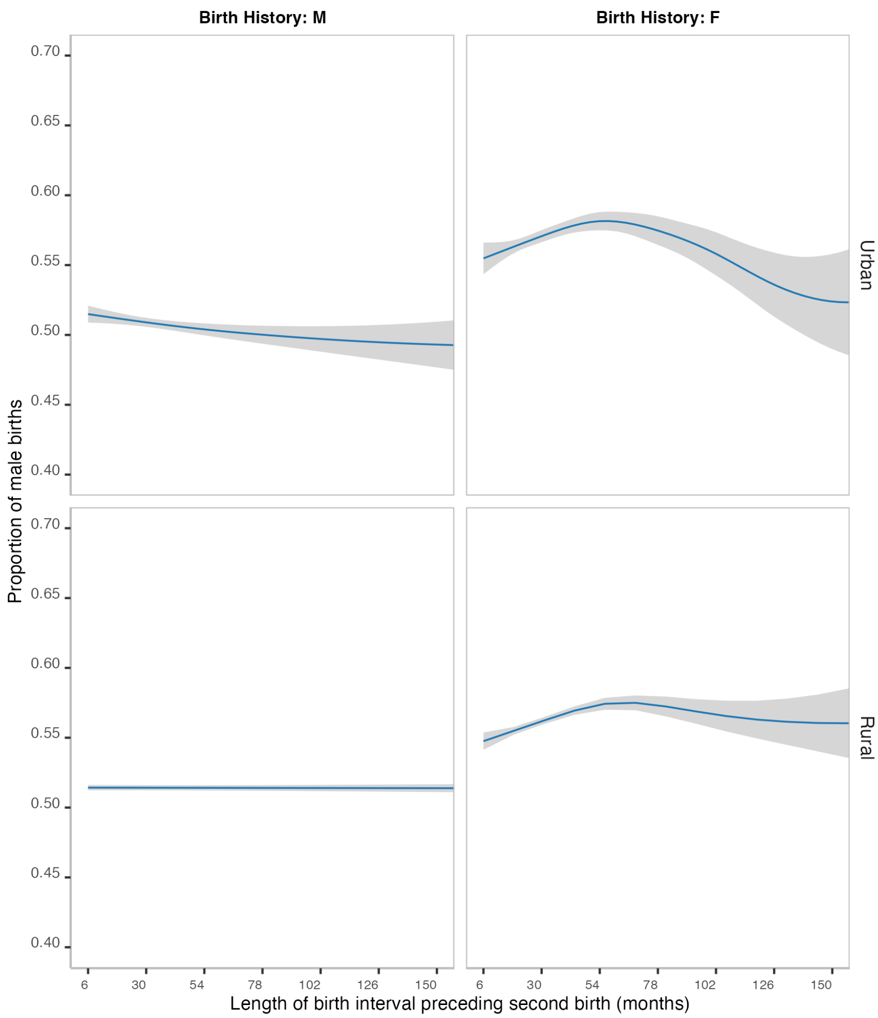
A. FIGURES

Figure A1 :
Variation in probability of male births (at the second birth order) at varying lengths of the preceding birth interval, by mother's education



Notes: This figure plots kernel-smoothed proportion of male births at the second birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories and across various levels of mother's education.

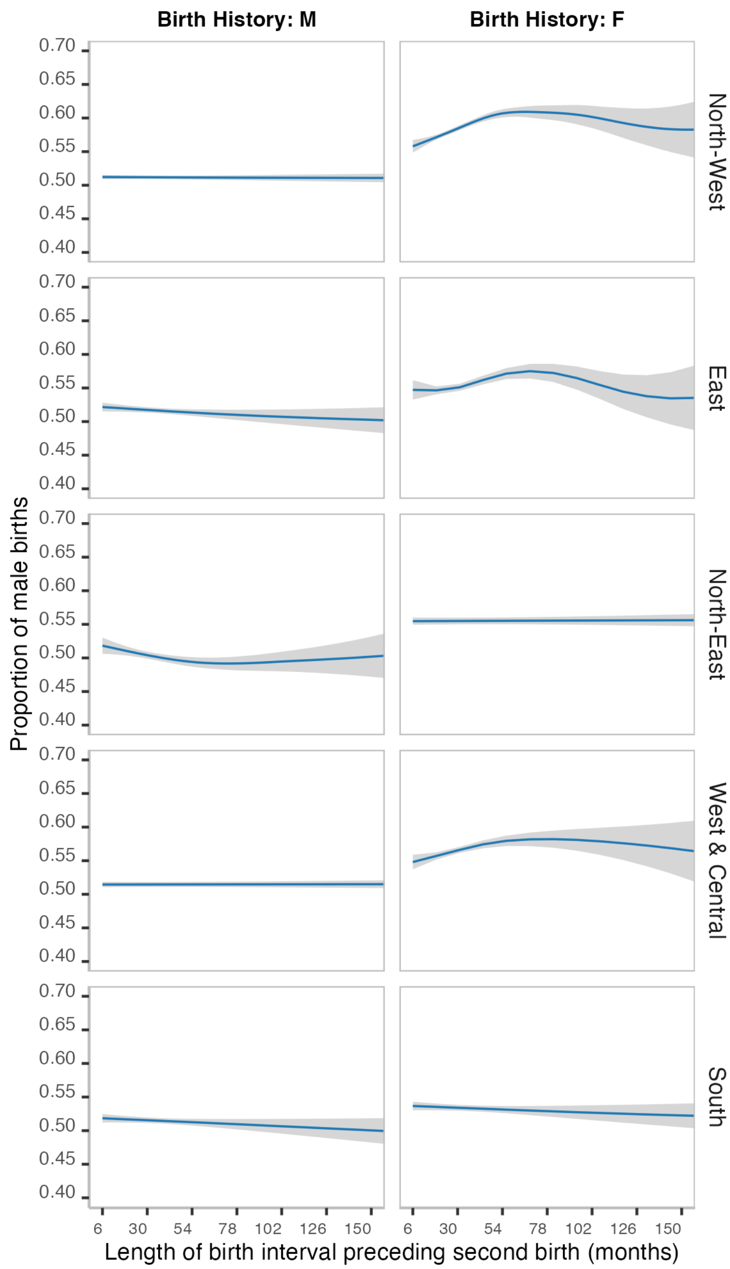
Figure A2 :
Variation in probability of male births (at the second birth order) at varying lengths of the preceding birth interval, by place of residence



Notes: This figure plots kernel-smoothed proportion of male births at the second birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories and across urban/rural settings.

Figure A3 :

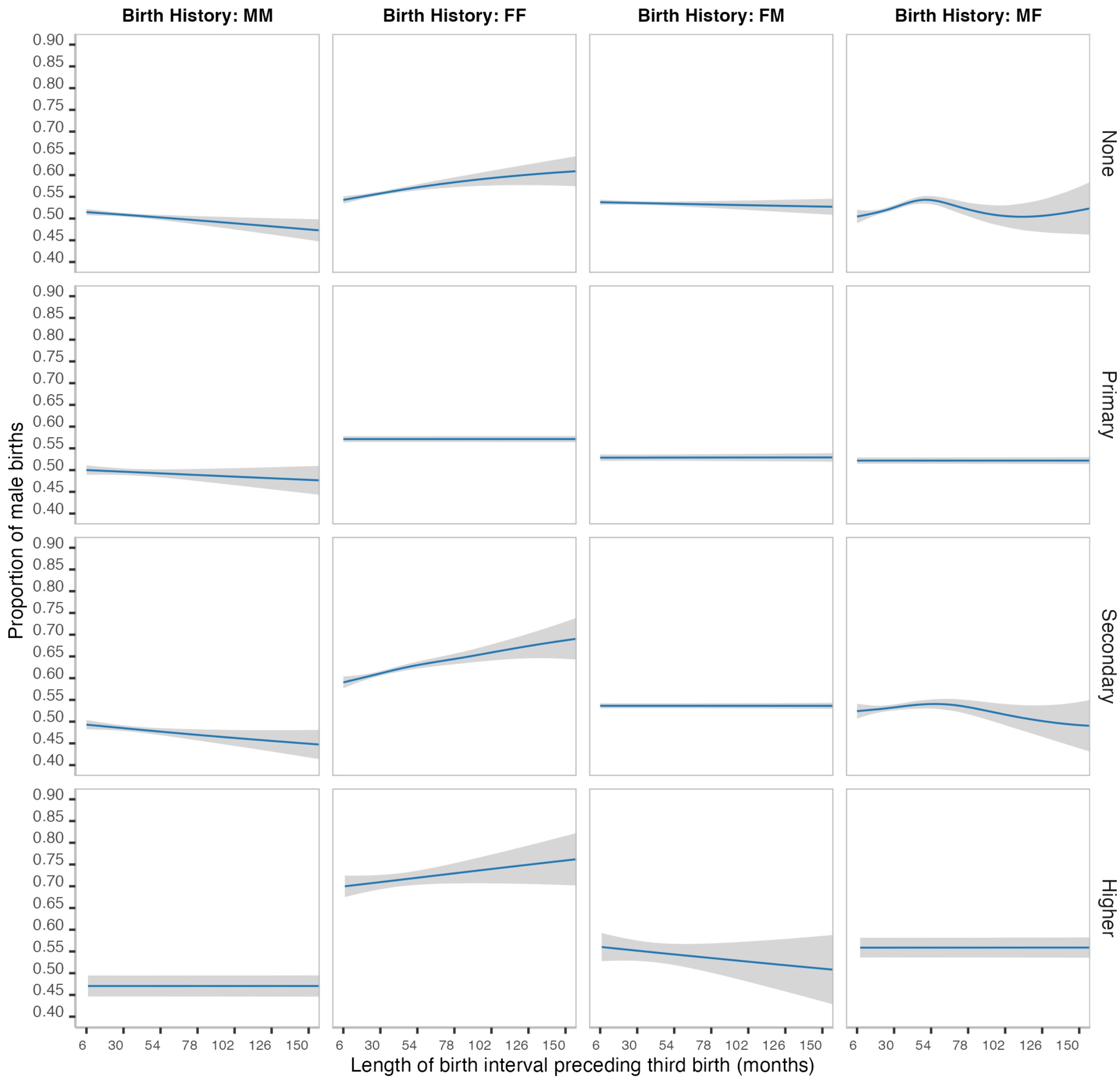
Variation in probability of male births (at the second birth order) at varying lengths of the preceding birth interval, by geographical region



Notes: This figure plots kernel-smoothed proportion of male births at the second birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories and across various geographical regions in India.

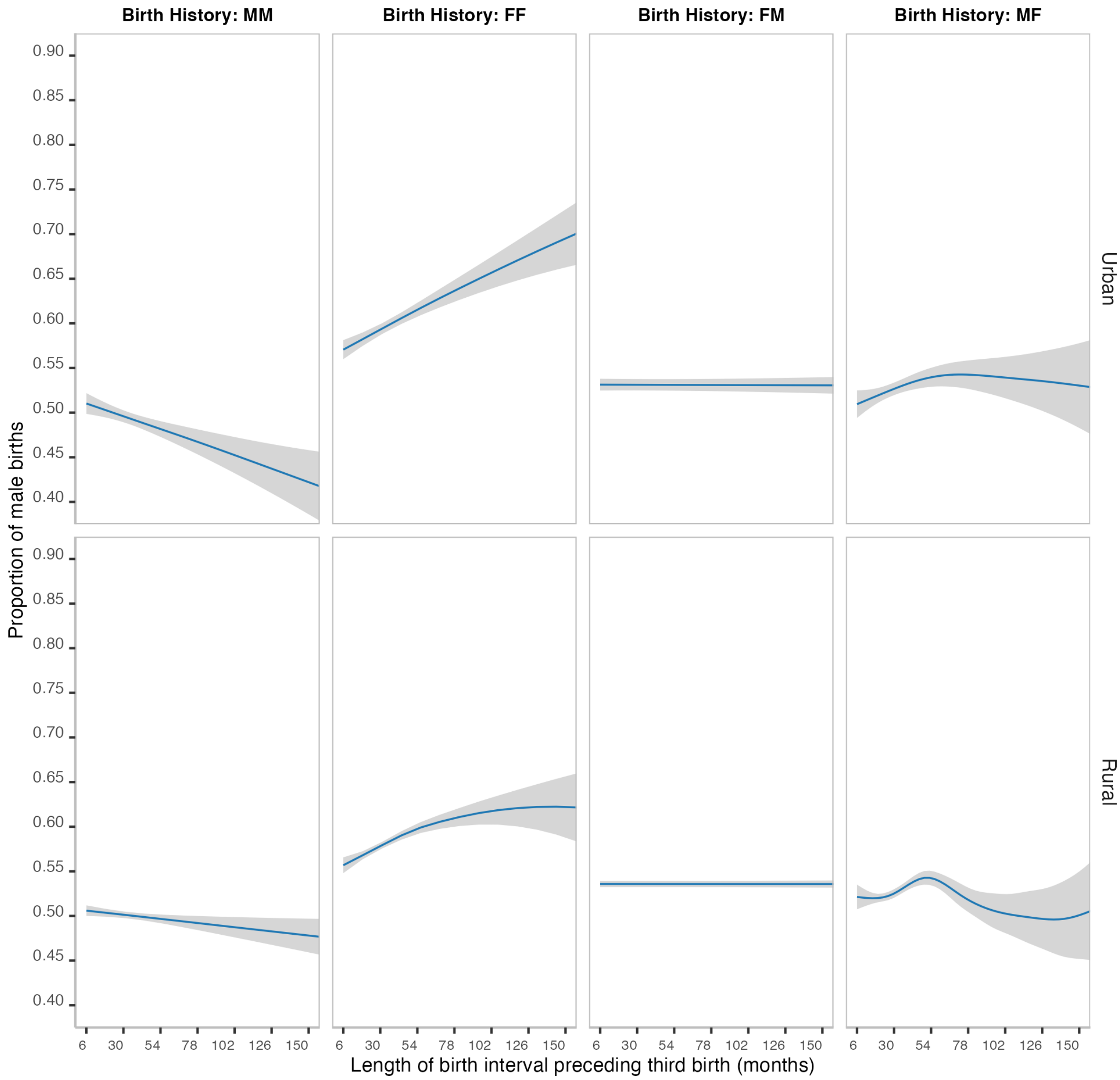
Figure A4 :

Variation in probability of male births (at the third birth order) at varying lengths of the preceding birth interval, by mother's education



Notes: This figure plots kernel-smoothed proportion of male births at the third birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories and across various levels of mother's education.

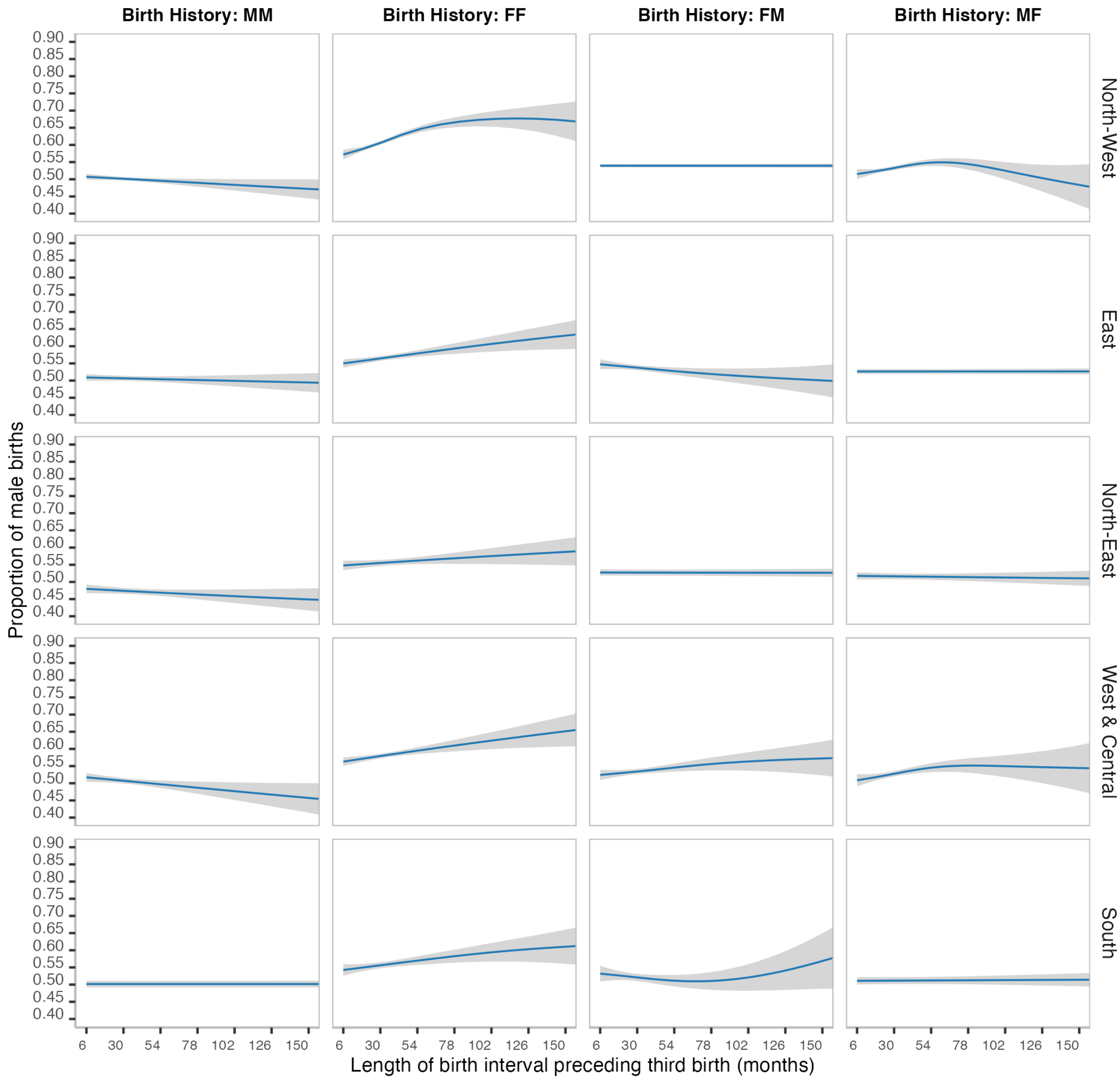
FIGURE A5 :
Variation in probability of male births (at the third birth order) at varying lengths of the preceding birth interval, by place of residence



Notes: This figure plots kernel-smoothed proportion of male births at the third birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories and across urban/rural settings.

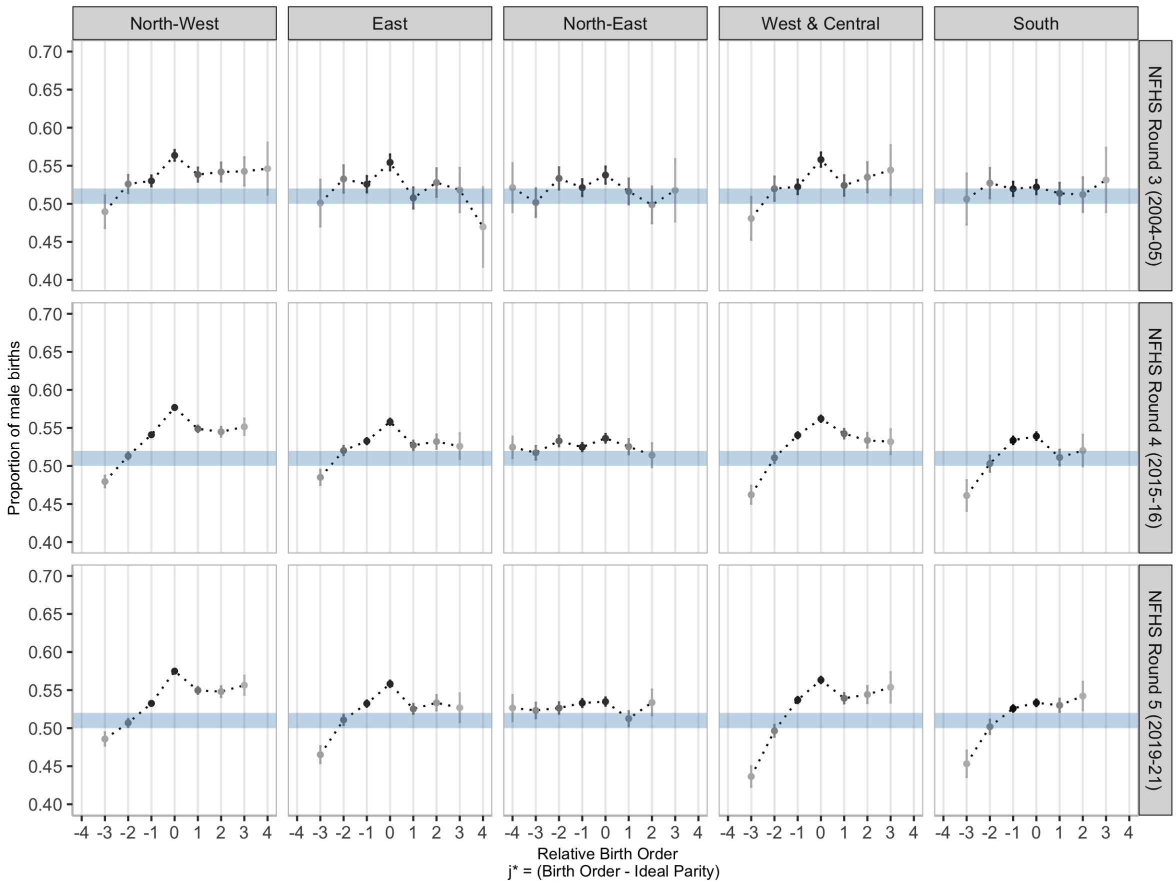
Figure A6 :

Variation in probability of male births (at the third birth order) at varying lengths of the preceding birth interval, by geographical region



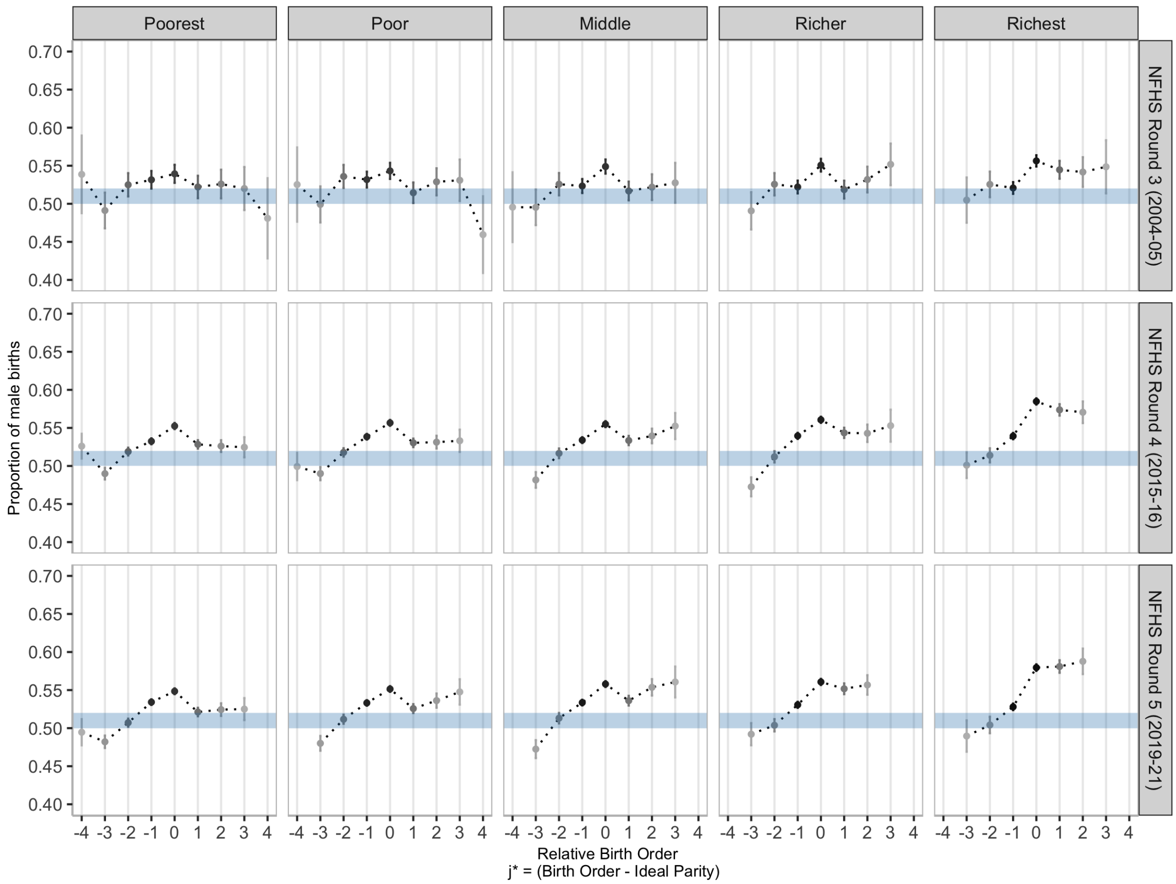
Notes: This figure plots kernel-smoothed proportion of male births at the third birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories and across various geographical regions in India.

Figure A7 :
Proportion of male births at relative birth orders, by geographical region



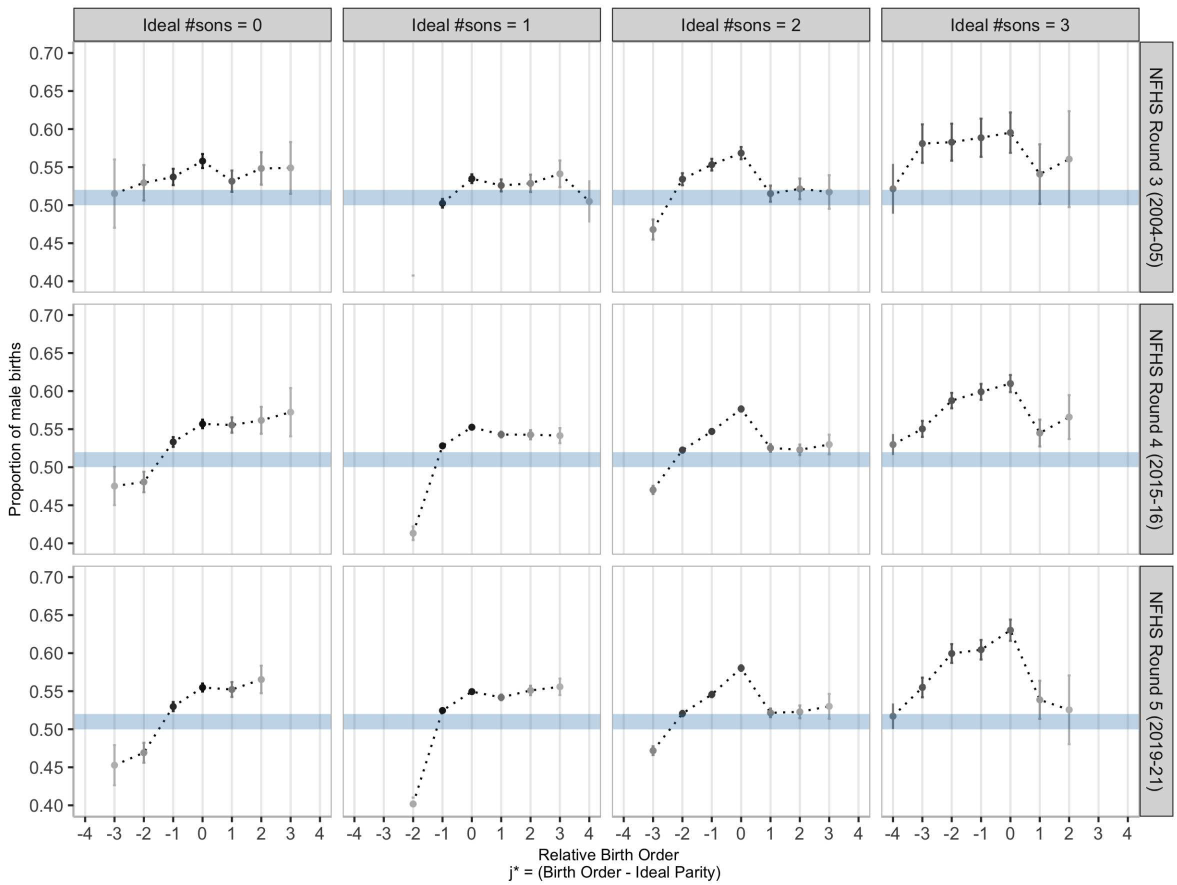
Notes: This figure plots the proportion of male births at various relative birth orders, by geographical regions. The vertical columns represent the geographical region and the horizontal panels specify the NFHS round. The shaded horizontal bar indicates the range 0.50-0.52 denoting the natural probability of male births Orzack et al. (2015). The color gradient of the markers indicate the share of births in the sample contributing to the statistic at the relative birth order, with darker shades representing higher shares.

Figure A8 :
Proportion of male births at relative birth orders, by geographical region



Notes: This figure plots the proportion of male births at various relative birth orders, by the parents' wealth quintile. The vertical columns represent the five wealth quintiles and the horizontal panels specify the NFHS round. The shaded horizontal bar indicates the range 0.50-0.52 denoting the natural probability of male births Orzack et al. (2015). The color gradient of the markers indicate the share of births in the sample contributing to the statistic at the relative birth order, with darker shades representing higher shares.

FIGURE A9 :
Proportion of male births at relative birth orders, by reported ideal number of sons



Notes: This figure plots the proportion of male births at various relative birth orders, by mother's reported ideal number of sons. The vertical columns represent the ideal number of sons reported and the horizontal panels specify the NFHS round. The shaded horizontal bar indicates the range 0.50-0.52 denoting the natural probability of male births Orzack et al. (2015). The color gradient of the markers indicate the share of births in the sample contributing to the statistic at the relative birth order, with darker shades representing higher shares.

B. TABLES

TABLE B1 :
Correlation between ideal and actual parity

	Dependent variable: Ideal parity		
	NFHS-3 (2005-06)	NFHS-4 (2015-16)	NFHS-5 (2019-21)
Total parity	0.290***	0.356***	0.362***
	(0.002)	(0.001)	(0.001)
$\mathbb{1}$(multiple births)	-0.019	0.064**	0.020
	(0.055)	(0.027)	(0.025)
Total parity $\times \mathbb{1}$(multiple births)	−0.040***	−0.064***	−0.059***
	(0.010)	(0.027)	(0.006)
Constant	1.455***	1.329***	1.315***
	(0.007)	(0.003)	(0.003)
Observations	67,441	371,865	379,827
R^2	0.262	0.256	0.256
Note:			*p<0.1; **p<0.05; ***p<0.01

Notes: This table examines the correlation between mothers' report ideal parities and their actual realized parity. In particular, the correlation is low and reduces in the case of multiple births i.e. when a pregnancy produces more than one baby (e.g. twins, triplets).

Table B2 :
Empirical validation of the heuristic at the fourth birth order

	Dependent variable: Interval preceding **fourth** birth		
	All mothers with $N \geq 4$	Mothers with $n^* = 4$ (Heuristic)	Mothers with $n^* > 4$ (Placebo)
	(1)	(2)	(3)
hist: FFF	-1.413^{***}	-1.259^{***}	-1.644^{**}
	(0.241)	(0.442)	(0.773)
hist: FFM	-0.908^{***}	0.039	-1.035
	(0.245)	(0.440)	(0.801)
hist: FMF	-1.745^{***}	-0.970^{**}	-2.498^{***}
	(0.250)	(0.450)	(0.782)
hist: FMM	-0.367	0.549	-1.281
	(0.272)	(0.471)	(0.828)
hist: MFF	-1.268^{***}	-0.372	-0.963
	(0.250)	(0.458)	(0.797)
hist: MFM	-0.642^{**}	0.302	-1.369^{*}
	(0.271)	(0.470)	(0.808)
hist: MMF	-0.486^{*}	0.266	-1.425^{*}
	(0.274)	(0.464)	(0.817)
sex: M	-0.708^{***}	0.376	-0.528
	(0.263)	(0.475)	(0.812)
hist: FFF × sex: M	1.387^{***}	0.042	0.179
	(0.329)	(0.619)	(1.077)
hist: FFM × sex: M	0.293	-1.029^{*}	0.739
	(0.335)	(0.613)	(1.116)
hist: FMF × sex: M	1.027^{***}	0.094	3.034^{***}
	(0.344)	(0.630)	(1.119)
hist: FMM × sex: M	0.624	-0.323	1.451
	(0.389)	(0.740)	(1.156)
hist: MFF × sex: M	0.895^{**}	-0.261	-0.894
	(0.349)	(0.632)	(1.098)
hist: MFM × sex: M	0.686^{*}	-0.700	0.518
	(0.376)	(0.676)	(1.132)
hist: MMF × sex: M	0.565	-0.314	2.058^{*}
	(0.378)	(0.675)	(1.147)
Constant	28.672^{***}	28.107^{***}	30.581^{***}
	(0.552)	(1.069)	(2.695)
Observations	214,529	62,176	16,017
Note:			*p<0.1; **p<0.05; ***p<0.01

This table reports the difference-in-difference estimates based on the specification in Equation (4) at the fourth birth order. Specification (1) reports the estimates for all mothers who give birth to at least four children. Specifications (2) and (3) report the estimates for mothers who according to our heuristic are expected to sex-select i.e. those whose ideal parity is equal to the given birth order ($n^* = 4$) and a placebo sample. The controls include mother's age at previous birth, rural, wealth quintile, NFHS round-specific fixed effects and state fixed effects. Robust standard errors included in parentheses.