Investigating the Structure of Son Bias in Armenia with Novel Measures of Individual Preferences*

Matthias Schief[1], Sonja Vogt[2,3], and Charles Efferson[4]

[1]Department of Economics, Brown University, U.S.A.
[2]Department of Social Sciences, University of Bern, Switzerland
[3]Centre for Development and Environment, University of Bern, Switzerland
[4]Faculty of Business and Economics, University of Lausanne, Switzerland

Last updated: March, 2021

Abstract

Sex ratios at birth favoring boys are being documented in a growing number of countries, a pattern indicating that families selectively abort females. Son bias also explains why in many countries girls have more siblings and are born at relatively earlier parities compared to their brothers. In this paper, we develop novel methods for measuring son bias using both questionnaire items and implicit association tests, and we collect data on fertility preferences and outcomes from 2700 participants in Armenia. We document highly skewed sex ratios, suggesting that selective abortions of females is widespread among parents in our sample. We also provide evidence that sex-selective abortions are under-reported, which highlights the problem of social desirability bias. We validate our methods and demonstrate that conducting implicit association tests can be a successful strategy for measuring the relative preference for sons and daughters when social desirability is a concern. We investigate the structure of son-biased fertility preferences within households, across families, and between different regions in Armenia, using measures of son bias at the level of the individual decision-maker. We find that men are on average considerably more son-biased than women. We also show that regional differences in son bias exist and that they appear unrelated to the socio-economic composition of the population. Finally, we estimate the degree of spousal correlation in son bias and discuss whether husbands are reliably more son-biased than their wives.

*Address correspondence to matthias_schief@brown.edu, sonja.vogt@soz.unibe.ch, and charles.efferson@unil.ch. We would like to thank UNICEF, Armenia, the Swiss National Committee of UNICEF, the Women's Resource Center (Yerevan), the Statistical Committee of the Republic of Armenia, and Ernst Fehr. SV and CE would like to thank the Swiss National Science Foundation (No. 100018_185417/1).

I Introduction

The most recent population census in Armenia counts 387,501 children below the age of nine. Among them 206,994 are boys but only 180,507 are girls. These numbers imply a sex ratio of over 114 boys for every 100 girls – a striking deviation from the natural sex ratio at birth of about 105 boys per 100 girls[1]. Skewed sex ratios have also been documented in Albania, Azerbaijan, China, Georgia, India, Korea, Montenegro, Pakistan, Tunisia, and Vietnam [United Nations Population Division, 2019, Chao et al., 2019].

Little is known, however, about the preferences that give rise to elevated sex ratios at birth. Many studies focus exclusively on fertility outcomes, which provide only limited information about the underlying fertility preferences. While fertility outcomes can reveal son bias at the group level, they are a very imprecise measure of fertility preferences at the household level and cannot be used to distinguish between the relative preferences of mothers and fathers within households[2]. Studies that do attempt to measure son-biased fertility preferences at the level of the individual often rely exclusively on survey data, which are likely compromised by social desirability bias.

Social desirability bias describes the tendency of individuals to answer questions in ways that will be viewed favorably by others, and social desirability constitutes one of the most important sources of bias affecting responses in survey-based research studies. Strong social norms shape whether parents are willing to reveal son-biased fertility preferences openly, and responses to explicit questions about the preferred gender of children will therefore likely be biased. In the context of Armenia, various non-governmental organizations have recently launched information campaigns targeted at reducing son bias in prospective parents, and the government has passed legislation to curb sex-selective abortions [International Center for Human Development, 2016, Sexuality Policy Watch, 2016]. As a consequence, parents in Armenia may have become less willing to talk openly about abortions and son-biased fertility preferences [Jilozian and Agadjanian, 2016, Fenton et al., 2001]. This complicates the study of son bias as it is unclear whether differences in stated fertility preferences across individuals or groups of individuals should be interpreted as stemming from differences in the degree of son bias, or whether they simply reflect differences in the severity of social desirability bias.

For this study, we designed novel implicit association tests and we administered these tests to a large sample of 2700 participants who were visited at their homes between May 2017 and March 2019. Our implicit association tests avoid asking participants directly about fertility preferences. Instead, we infer son bias from reaction times in categorization tasks involving drawings of families with sons only, drawings of families with daughters only, and value-laden words. We also build on recent work by Jayachandran [2017] to complement our implicit association tests with a measure of explicit son bias based on survey questions, which overcomes some important biases of similar measures that have been used in the literature. In this paper, we evaluate the validity of these measures and we use our

[1]The numbers reported here are based on the 2011 population census. A new population census is currently underway, but the data are not yet publicly available at the time of writing. More recent data from Armenia's Civil Status Register report 124,343 live births of boys and 111,358 live births of girls over the period 2014–2019, suggesting that the sex ratio at birth currently stands at around 112 boys for every 100 girls – a moderate reduction from earlier levels but still much higher than the natural sex ratio at birth [The Demographic Handbook of Armenia, 2020].

[2]Similarly, birth records alone cannot distinguish between the role of parental demand for sex selection and the supply of medical ultrasound – a necessary precondition for sex-selective abortions. This problem is recognized, for example, by researchers who study whether the surge in the sex ratio at birth that has been observed in the Caucasus following the dissolution of the Soviet Union is best understood in terms of changing fertility preferences on the parts of parents or the increased availability of ultrasonography [Duthé et al., 2012, Hohmann et al., 2014].

data to generate important insights about the structure of son bias within and across households in Armenia.

Our paper contributes to the literature on measuring fertility preferences and to the literature on distorted sex ratios. We contribute to the literature on measuring fertility preferences by implementing and validating novel measures of son bias at the level of the individual decision-maker [Coombs, 1974, McClelland, 1979, 1983, Bongaarts, 1990, Fuse, 2010, Norling, 2018]. Because we measure fertility preferences of both mothers and fathers, we can use our data to study spousal differences in son bias [Lundberg, 2005, Dahl and Moretti, 2008, Robitaille, 2013, Robitaille and Chatterjee, 2018, Hassan et al., 2019]. Our methodology for measuring implicit son bias also relates to recent papers that have used implicit association tests in the field to elicit attitudes on sensitive topics [Beaman et al., 2009, Lowes et al., 2015, Efferson et al., 2015, Vogt et al., 2016]. Furthermore, our paper relates to research on the elevated sex ratio in the Caucasus by directly measuring the fertility preferences of parents who selectively abort female fetuses [Meslé et al., 2007, Duthé et al., 2012, Michael et al., 2013, Guilmoto, 2013, Guilmoto and Tove, 2015]. Finally, by providing measures of preferences, we hope that our paper can also be informative for more theoretical approaches that study the interplay of preferences with technology diffusion and fertility decline [Kashyap and Villavicencio, 2016].

The paper is structured as follows. In section II, we show that our data provide clear evidence for both the prevalence of sex selection in our sample, as well as for significant underreporting of sex-selective abortions. This underscores the problem of social desirability and motivates the development of novel measures of son bias. Section III introduces our measures of implicit and explicit son bias and validates them. In section IV, we use our measures to study the structure of son bias in Armenia, and section V concludes.

II Prevalence and underreporting of sex-selective abortions

A straightforward way to study son bias and sex selection in Armenia is to ask parents directly about sex-selective abortions or their relative preference for sons over daughters. In our questionnaire, we included a module that asked female participants about their reproductive histories, including miscarriages and abortions at different gestational ages, and whether the sex of the fetus was known at the time of the abortion. Relying exclusively on self-reports, however, increases the risk of drawing wrong conclusions from data that suffer from social desirability bias. While this problem is generally well recognized, overcoming social desirability bias has remained a challenge. The standard approach to dealing with social desirability is to provide anonymity and privacy to the study participants, so that their responses will be less affected by reputation concerns.

To test the feasibility of this approach, we adopted a research protocol that emphasized the high degree of privacy that we were able to provide for our participants[3]. We physically separated the household members for the purpose of data collection by interviewing them in different rooms of their home, so that there was no opportunity for family members to influence each others' responses[4].

[3]High degrees of privacy are standard in research studies that can be conducted in the lab. Data collection in the field typically involves compromises that result in lower degrees of privacy. For example, participants in the *Demographic and Health Surveys (DHS)* and the *Multiple Indicator Cluster Survey (MICS)* – the two most prominent surveys that collect data about parental gender preferences – are interviewed in a face-to-face setting and responses to sensitive questionnaire items must be communicated to the facilitators who record them.

[4]In some household there were not enough rooms to assign each household member to a different room. In this case, our facilitators separated the participants with the help of mobile room dividers.

To make the physical separation salient, husbands and wives were interviewed at the same time by different facilitators. The most sensitive questions, including any question about abortions, were asked in a computerized self-administered questionnaire that allowed our participants to record their answers in a completely private setting. The facilitators were instructed to leave the room while the participants answered these questions, and the responses were not accessible for the facilitators after the participants had completed the self-administered questionnaire. Finally, the data collection was overseen by an independent monitoring team, the Women's Resource Center of Armenia, which ensured a strict adherence to the privacy protocol.

We collected data in three regions of Armenia, namely the capital region of Yerevan, as well as Gegharkunik and Syunik, the regions with the highest and the lowest sex ratio among children in the 2011 census[5]. Within each region, we randomly sampled married couples with at least one child under 16 living at home according to a two-stage sampling frame described in detail in appendix A.1. Whenever possible, we collected data from the husband, the wife, and the husband's mother[6]. We now present evidence for sex selection in our data and show that, despite our extensive efforts to minimize the risk of social desirability bias, we have clear indications that sex-selective abortions are under-reported. This result underscores the problem of social desirability and motivates the use of implicit association tests to complement the survey data.

II.1 Evidence for sex selection

The sex ratio among children in our sample is 112.8 boys for every 100 girls. This ratio is consistent with population-level sex ratios in Armenia computed from birth registry data [Guilmoto, 2013][7]. While preferentially choosing to stop having children after the birth of a son can affect the sex ratio within families, an elevated sex ratio *across families* can only result from parents actively engaging in sex selection. In the context of Armenia this typically takes the form of sex-selective abortions[8].

Figure 1 plots sex ratios among first-, second-, and third-born children in our sample[9]. These disaggregated results reveal a striking pattern. While the sex ratios among first- and second-born children are statistically indistinguishable from the natural ratio of 105 boys for every 100 girls, the sex ratio among third-born children is significantly higher at about 177:100. This finding is consistent with previous research documenting an elevated sex ratio among third-born children in Armenia and attests to a widespread practice of sex selection in favor of boys [Meslé et al., 2007, Guilmoto, 2013].

We further disaggregate the sex ratio by focusing on children born to parents who still lacked a son or lacked a daughter[10]. Interestingly, we find that the spike in the sex ratio among third-born children

[5]Specifically, we focus on the sex ratio among children below the age of 16. The ratio of boys to girls is 1.08 in Syunik, 1.22 in Gegharkunik, and 1.12 in Yerevan.

[6]Our motivation for collecting data from a husband's mother stems from the fact that young couples in Armenia often share their household with the husband's parents. In conversations with our implementation partners, we learned that the husband's mother has a reputation for exerting influence on the fertility decisions taken by young couples. This prompted us to collect data not only from the focal couple but also from the husband's mother. The focus of this paper, however, is not on the role of mothers-in-law in shaping fertility outcomes, and we leave this question for future research.

[7]See appendix A.1 for a more detailed discussion of the representativeness of our sample.

[8]Section III.3 discusses evidence of son-biased fertility stopping behavior in our data.

[9]We restrict our attention to the sex ratios for first-, second-, and third-borns, since only 6% of all sampled households have 4 or more children. See figure A.1 in appendix A.1 for the sex ratios among children born fourth and later. Importantly, to compute the sex ratios in figure 1 we use data from all households, including households with 4 or more children.

[10]See Guilmoto [2017] for a recent study that follows a similar strategy to study son-biased fertility outcomes in Georgia, Indonesia, and Vietnam.

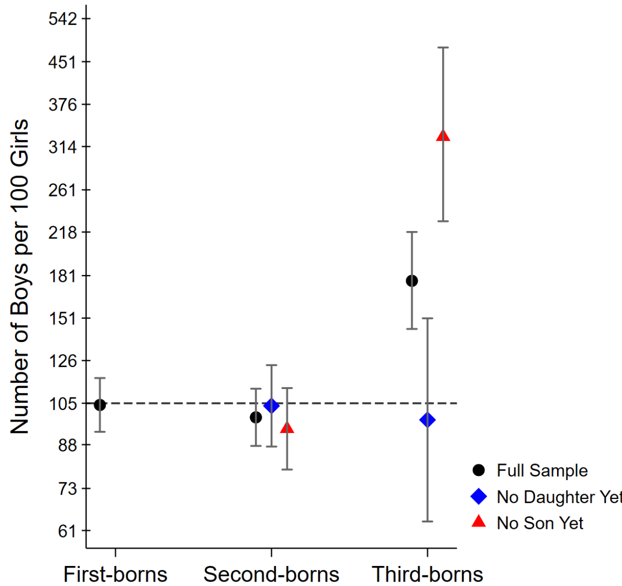

Figure 1: The ratio of boys to girls in our sample, computed separately by birth order. 95% confidence intervals are computed using the Clopper-Pearson method for calculating binomial proportion confidence intervals.

is particularly pronounced for families in which the first two children are girls. In these families, the sex ratio among third-born children reaches a staggering 328 boys for every 100 girls. Tellingly, when we condition on families without daughters, the sex ratio of third-borns remains indistinguishable from the natural ratio, suggesting that it is indeed the absence of a son rather than the desire for a more balanced gender composition that induces parents in Armenia to engage in sex selection[11].

II.2 Evidence for under-reporting of sex-selective abortions

To test for under-reporting, we compute the number of missing girls in our sample and compare it against the number of reported abortions. We followed a stepwise procedure when recording abortion histories. We began by asking the participants whether they ever had an abortion[12]. Then, if this question was answered affirmatively, we proceeded to ask the participants about abortions after the eighth week of pregnancy[13]. If the participant reported having had abortions after the eighth week of pregnancy, we asked about the timing of these abortions. In particular, for each abortion we asked if it had taken place before the birth of the first child, after the birth of the first but before the birth of

[11] In appendix A.2, we show that we obtain qualitatively very similar results if we conduct this analysis for each region separately (figure A.2) and we further dis-aggregate the sex ratio among third-born children by plotting it separately for all possible gender compositions among the first two children (figure A.3).

[12] Answer categories are *yes*, *no*, and *prefer not to answer*. Only 3.57% of all respondents preferred not to answer.

[13] The sex of a fetus can be reliably ascertained via ultrasounds starting from the 12^{th} week of pregnancy [Mazza et al., 2001]. Analysis of fetal DNA found in the mother's blood allow for sex discernment at an earlier moment in the first trimester, but such analyses are not carried out routinely. We want to distinguish potentially sex-selective abortions from all other abortions that were carried out earlier. However, given that abortions beyond the 12^{th} week of pregnancy are illegal in Armenia, except for special medical circumstances, we feared that asking respondents about abortions after the 12^{th} week would result in under-reporting. Thus, we asked respondents about abortions after the eighth week of pregnancy.

the second child, etc. Finally, we asked whether the participant knew if she was expecting a boy or a girl.

Abortion rates are high in Armenia, a fact that reflects the country's history of reliance on induced abortions as a method of fertility regulation [Remennick, 1991, David, 1992]. Consistently, we find that about one third of all mothers in our sample report having had at least one abortion, and many report having had multiple abortions. Among those who report any abortions, about two thirds report having had at least one abortion after the 8^{th} week of pregnancy, and about one fourth report having had multiple such abortions between two births. While there may be little stigma associated with having had abortions to limit the number of children, it is unclear whether participants are willing to reveal that they are selectively aborting female fetuses.

Because the sex ratio is only distorted at higher parities (Fig. 1), we separately compute the number of missing girls among first-, second-, and third-born children[14]. To compute the number of missing girls, we count the number of sons in our sample and based on this derive the expected number of daughters implied by a natural sex ratio of 105 boys for every 100 girls [Sen, 1992]. The difference between the expected and the actual number of daughters constitutes our estimate of how many girls are missing in our sample. Note that this is a conservative estimate of the number of missing girls because it abstracts from the possibility of selective abortions of male fetuses.

Making use of the fact that our survey asks about the timing of the abortions, we can compare the number of missing girls among first-, second-, and third-born children to the number of reported abortions in the relevant interbirth interval. For example, the sex ratio among second-born children will be affected by sex-selective abortions after the birth of the first, but before the birth of a second child. Similarly, sex-selective abortions after the birth of the second child, but before the birth of a third child matter for the sex ratio among third-born children. Our focus is on the question of whether there are enough reported abortions in the relevant interbirth interval to account for the missing girls in our sample. More specifically, as a telltale sign of underreporting, we will look to see if the number of reported abortions in our sample falls below the lower bound of the 95% confidence intervals associated with our estimates of missing girls.

Panel A of figure 2 plots the number of missing girls in our full sample along with the number of self-reported abortions after the eighth week of pregnancy. We estimate that, among third-born children in our sample, there are 102 girls missing. At the same time, our participants report about 250 abortions that occurred after the birth of the second and before the birth of a third child. Hence, we observe more than enough self-reported abortions to account for the missing girls, and we cannot reject the null hypothesis that there is no underreporting of abortions. This follows from the fact that the number of reported abortions in general, without conditioning on the sex of the fetus, does not fall below the confidence intervals for missing girls. Following this logic, however, we do find unambiguous evidence for under-reporting of *sex-selective* abortions. In particular, the number of self-reported abortions where the mother reports knowing that she was expecting a girl is strictly below the 95% confidence interval for missing girls among third-born children.

In panel B of figure 2, we repeat this exercise on the restricted sample, where we exclude self-reported abortions and births that occurred after a family already had a son. Two findings stand out. First, we find that almost all missing girls can be linked to parents who did not yet have a son. This is consistent with figure 1 and can be seen by the fact that the total number of missing girls among

[14] While two children are the most common family size in our sample, about one third of all sampled households have a third child.

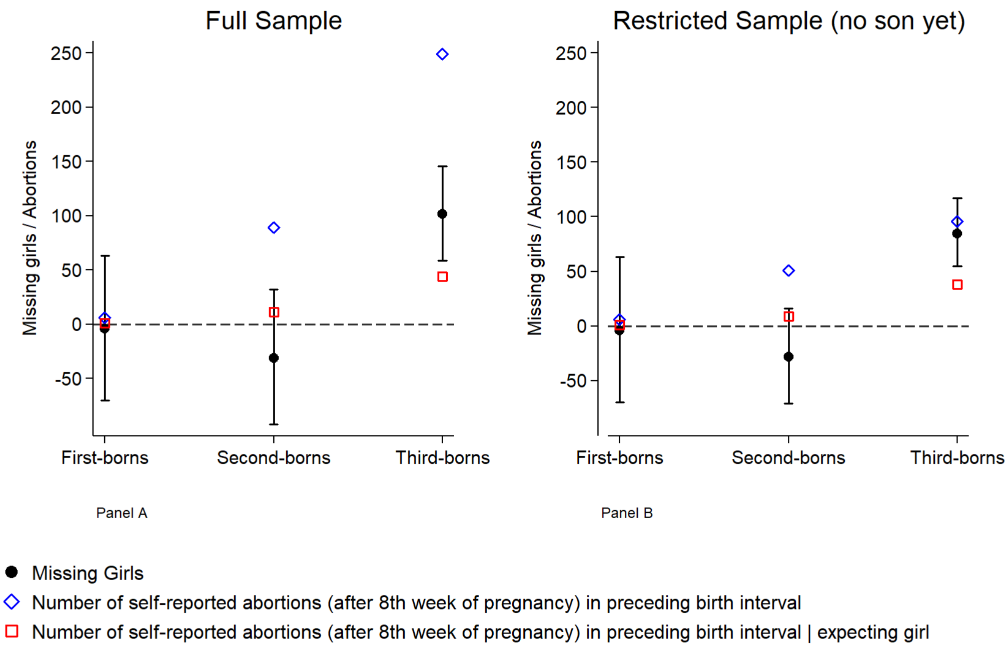

Figure 2: Evidence for under-reporting of sex-selective abortions. We infer the *expected* number of girls from the observed number of boys in our sample by modeling the gender of a child as an i.i.d. draw from a Bernoulli distribution and assuming a natural sex ratio at birth of 105 boys for every 100 girls. Our estimate of missing girls is simply the difference between the actual and the expected number of girls in our sample. We separately count the number of self-reported abortions preceding the birth of first-, second-, and third-born children. Sex-selective abortions occurring after the birth of a second child but before the birth of a third child affect the sex ratio among third-born children, even if no third child is ever born to the mother who has the abortion. Hence, the preceding birth interval for third-born children begins after the birth of the second child and, in the event of a third birth, ends before that birth. Similarly for second-born and first-born children.

third born children is almost as high in panel B as it is in panel A. Second, not only is the number of abortions where the mother reports knowing that she was expecting a girl again too low to account for the number of missing girls among third-born children, but the overall number of self-reported abortions is itself only slightly higher than the number of missing girls. This implies that almost all of the self-reported abortions that occurred after the birth of second children in this sample must have in fact been sex-selective, even though in the majority of cases the respondents claim that they did not know whether they were expecting a girl or a boy[15].

To summarize, we find that many participants falsely report that they did not know the sex of the fetus when deciding on whether to end the pregnancy – presumably in order to conceal the sex-selective nature of the abortion. We take these findings as a stark reminder that, even with a research design

[15]Similarly, we can derive an estimate of the share of sex-selective abortions in the total number of reported abortions across all parities by comparing our estimate of the total number of missing girls to the total number of reported abortions in our sample. Ignoring the possibility of sex-selective abortions in favor of girls, we find that about 20% of all reported abortions that occurred after the eighth week of pregnancy must have been sex-selective in favor of boys. Hence, while sex selection appears to be the motive behind a significant share of abortions in Armenia, this statistic also shows that more abortions are carried out for reasons that are unrelated to son-biased fertility preferences.

like ours, which guarantees an unusual degree of privacy to the participants, untruthful reporting and social desirability bias remain a serious concern. Our primary objective in this paper is to study the structure of son-biased fertility preferences. But given the documented underreporting of sex-selective abortions, we must recognize the risk that social desirability may also induce participants to understate their degree of son bias when asked about it directly. An important advantage of our research design is that we complement measures of explicit son bias with implicit association tests to deal with the problem of social desirability. The next section will discuss our measures of son bias in more detail.

III Measuring son bias

We pursued two complementary strategies to measure an individual's degree of son bias. First, we developed and implemented implicit association tests that we specifically designed to measure son bias among Armenian participants. We will call this our measure of implicit son bias. Implicit association tests present participants with visual and auditory stimuli and require the participants to quickly sort these stimuli according to changing categorization schemes, thereby measuring a person's association between mental representations. Implicit association tests have been successfully used to elicit attitudes on sensitive topics like racism and gender norms, and are credited with overcoming social desirability bias [Greenwald et al., 1998, Asendorpf et al., 2002, Egloff and Schmukle, 2002, Kim, 2003, Nosek et al., 2005, Greenwald et al., 2009; but see also Karpinski and Hilton, 2001, Nosek et al., 2007][16]. Second, we built on Jayachandran [2017] to construct a survey-based measure of son bias at the level of the individual, which overcomes important biases characterizing past research. We will call this our measure of explicit son bias.

We will validate our methods by demonstrating the internal consistency of our measures of implicit and explicit son bias at the individual level, and by showing that our measures also correlate with observed fertility outcomes at both the household and community levels. Having shown that our implicit association tests are successful in measuring son bias, we will then use our results to establish important new insights about the structure of son bias in Armenia. In particular, we will show that men tend to be much more son-biased than women. Moreover, we will show that regional differences in son bias exist and that they appear unrelated to the socio-economic composition of the population. Finally, we will estimate the degree of spousal correlation in son bias and discuss whether husbands are reliably more son-biased than their wives.

III.1 Implicit association tests

We designed two versions of an implicit association test (IAT) to measure son bias, a *Valence IAT* and a *Stereotype IAT*. In both versions our participants were presented with drawings of two different Armenian families and audio recordings of value-laden words. The first family has two sons and the second family has two daughters. Both families were shown in typical everyday scenes (see appendix A.3 for examples of the drawings).

In the *Valence IAT*, each drawing was followed by an audio recordings of either a positive word (e.g. joyous, excellent, spectacular) or a negative word (e.g. sad, to fail, yucky). In the *Stereotype*

[16] Randomized response represents another well-known strategy for dealing with social desirability in studies about controversial attitudes or behavior. The principal idea of randomized response measures is to allow any individual participant plausible deniability. While randomized response techniques are useful tools to induce truthful responses and to elicit typical preferences at the group level, they cannot provide reliable measures of individual preferences.

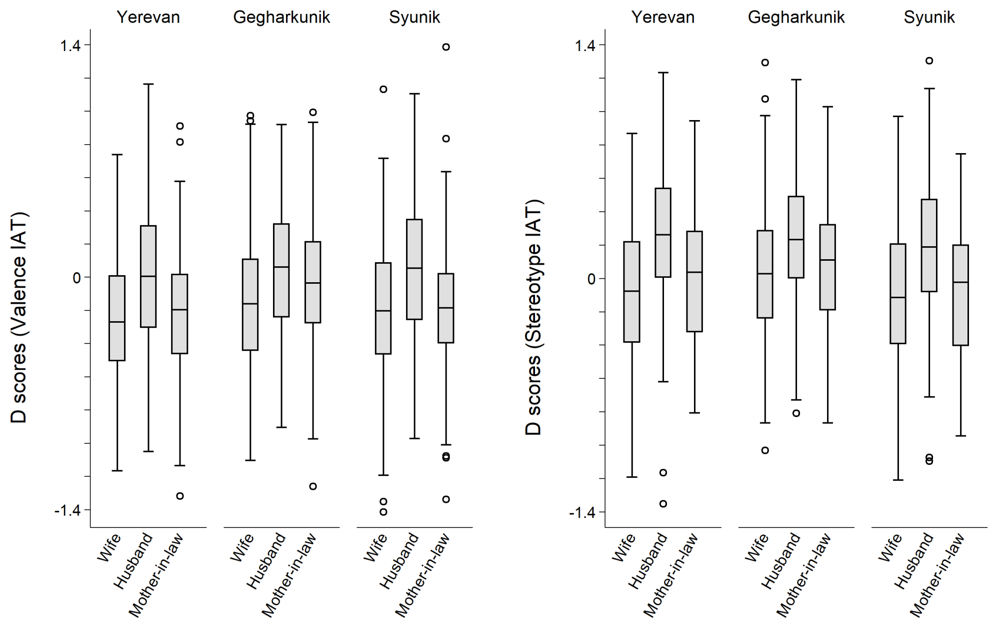

Figure 3: Distribution of D scores by region and subject type. The upper and lower ends of each box indicate the upper and lower quartiles, respectively, with the line inside the boxes indicating the median. The whiskers indicate the upper and lower adjacent values, and the circles indicate any observation that falls above or below the adjacent values.

IAT, we used the same drawings but selected words that pertain to stereotypes that associate sons with flourishing families. In particular, we used words belonging either to the semantic category "*flourishing*" (e.g. to multiply, to immortalize, descendant) or the semantic category "*withering*" (e.g. to fade, extinction, infertility).

In both implicit association tests, the participants had to quickly sort drawing and words according to changing categorization schemes that grouped the family with sons together with positive (flourishing) words and the family with girls together with negative (withering) words, or vice versa. Our measure of interest is the individual's D score, which is constructed from a participant's relative response time under the two sorting rules [Greenwald et al., 2003]. D scores are distributed on the interval $[-2, 2]$, where a positive score indicates an automatic mental association between positive (flourishing) words and sons, and between negative (withering) words and daughters.

Figure 3 summarizes the distribution of D scores by subject type and region. As we will show more formally later, it can already be seen here that men are more son-biased on average than women. This pattern is found in all regions and holds for both versions of the implicit association test. Households were randomly assigned to take either the Valence IAT or the Stereotype IAT. Because the two versions of the implicit association test produce similar results, we combine the D scores from both versions in most of the analyses in the paper. In the remainder of the paper, we will refer to the D scores as our measure of implicit son bias. For details about the design of the implicit association tests, the exact protocol for administering the tests, and the algorithm to compute D scores, we refer the reader to appendix A.3.

III.2 Survey-based measures of son bias

The typical approach for elicitation of a participant's ideal family composition involves asking the participant the following question[17].

> *If you could go back to the time you did not have any children and could choose exactly the number of children to have in your whole life, how many would that be?*

This question is then followed up by a second question.

> *How many of these children would you like to be boys, how many would you like to be girls, and for how many would the sex not matter?*

This methodology involves a number of risks. Most importantly, responses will suffer from rationalization bias as parents feel the need to report an ideal family size and gender composition that resembles their actual family [Westoff and Ryder, 1977, Rosenzweig and Wolpin, 1993]. Moreover, forcing respondents to condition their responses to the second question on their response to the first question can introduce additional bias. Consider a person who likes a balanced gender mix but would always choose a family composition with more sons than daughters over a family composition with more daughters than sons. We will call such a person weakly son-biased. Given the sequence of questioning above, however, this person's weak son bias will not be revealed if her optimal family family size corresponds to an even number of children. This conditioning bias becomes particularly problematic when comparing individuals across regions or socio-economic backgrounds who will differ in their preferred family sizes[18].

To overcome these biases, we built on Jayachandran [2017] to construct an explicit measure of individual son bias that is based on the following three principles. First, we asked each participant to specify the ideal gender composition for the future families of the youngest child in the household. While respondents should care about the fertility outcomes of their offspring, these hypothetical families did not exist at the time of the study, and so rationalization cannot affect responses. Second, we explicitly controlled for conditioning on family size by asking respondents to report their ideal gender composition for families of various sizes. This allowed us to separate preferences related to family composition from preferences related to family size. To avoid that anchoring confounds our results, we distributed these questions across different parts of the questionnaire and we randomized the order in which they were presented across participants. Third, we did not simply elicit the ideal sex ratio conditional on families of various sizes. Rather, we elicited the entire ideal ordering of sons and daughters.

To summarize the explicit fertility preferences of a given participant in a single statistic, we construct an index that captures the degree to which each participant's preferred family composition is biased toward sons[19]. The index is based on a scoring algorithm that considers whether the participant

[17] See, for example, in the *Demographic and Health Surveys*.

[18] A third weakness of the traditional approach for eliciting ideal family compositions is its inability to isolate preferences over the optimal number of children from preferences over the sex composition of children. For example, parents in many societies are concerned with having at least on male heir. When asked about the ideal family size, it is then unclear whether participants value large families per se or whether they see large families as a means to increase the likelihood of having at least one son.

[19] While Jayachandran [2017] implements a cross-subject design, we can construct such an index, because we ask each participant to specify the ideal gender composition for families of different sizes. We preregistered this index in Efferson et al. [2018].

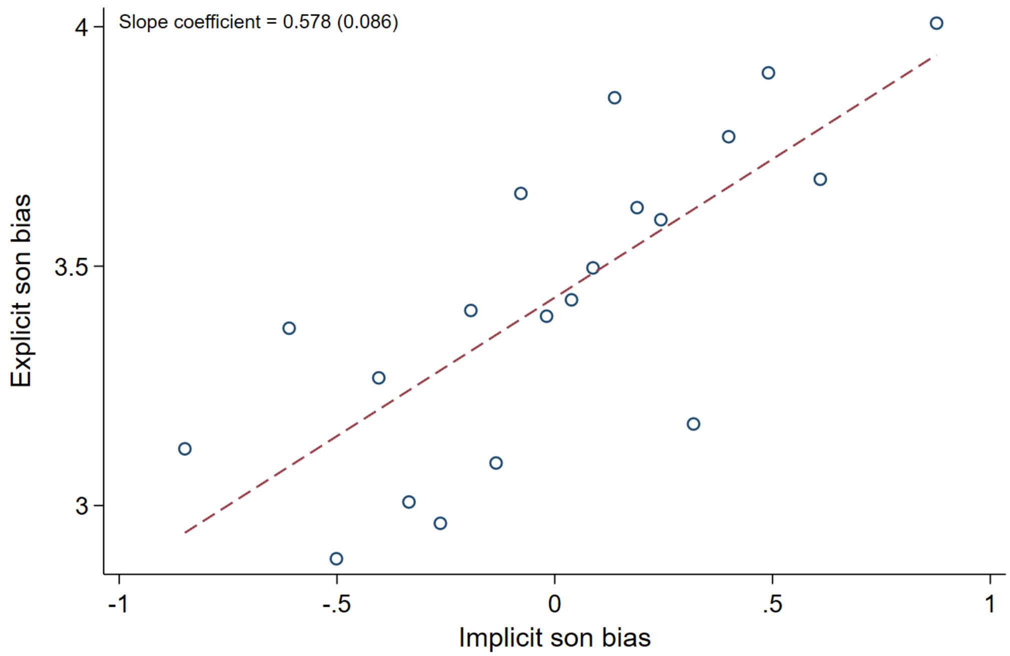

Figure 4: Binned scatter plot showing the individual-level association of our measures of implicit son bias (D scores) and explicit son bias (index). To decide on the optimal number of bins, we follow Cattaneo et al. [2019] who develop a data-driven decision procedure based on an integrated mean square error approximation.

(i) prefers a first child to be a son, (ii) prefers more sons than daughters, or (iii) prefers families with exclusively male children. The resulting index can take integer values between zero and nine. In the remainder of the paper, we will refer to this index as our measure of explicit son bias. A detailed description of our scoring algorithm and a histogram of the resulting index values are in appendix A.4.

III.3 Validating our measures of implicit and explicit son bias

To validate our novel measures of son bias, we subject them to two tests. First, we ask if our measures of implicit and explicit son bias are internally consistent at the individual level. Second, we ask if both measures of son bias are consistent with realized fertility outcomes[20].

Figure 4 shows the association between our measures of implicit and explicit son bias in a binned scatter plot. We find a positive and highly significant relationship between our two measures of son bias[21]. In appendix A.5 we demonstrate that this association remains positive and significant after accounting for the gender of the respondent, the gender composition of the children in the household, and the community in which the respondent resides.

Our measure of explicit son bias directly quantifies the degree of bias in a respondent's stated fertility preferences, and its key advantage is the fact that interpreting the measure is straightforward. The weakness of any explicit approach to measuring son bias is that one cannot rule out social desirability bias in the responses. Implicit association tests are useful in that they solve the problem of social

[20] In section A.5 of the appendix, we conduct a third validation test, in which we correlated our measures of son bias with a variable that approximates the question asked in the *Demographic and Health Surveys*.

[21] The raw correlation between our measures of implicit and explicit son bias is 0.128.

desirability bias by design. Their weakness is that an interpretation of D scores as measures of son bias hinges on the existence of a link between son-biased fertility preferences and relative response times across different categorization tasks. The positive association between the two measures of son bias documented in figure 4 reassures us that social desirability bias does not render the index uninformative, and that our implicit association tests indeed capture variation across participants that is closely related to their fertility preferences. We interpret the internal consistency of our two measures as evidence that both D scores and index values are useful proxies of latent son bias.

In a second validation exercise, we assess the behavioral relevance of our measures by asking whether higher levels of measured son bias are positively associated with son-biased fertility outcomes. Son-biased parents tend to keep having children until they reach their desired number of sons, and it is well understood that this fertility strategy results in girls on average having a greater number of siblings and boys having a higher chance of being the last-born child [Yamaguchi, 1989, Clark, 2000, Filmer et al., 2009, Basu and De Jong, 2010, Altindag, 2016]. It also increases the expected share of boys in each household[22]. Our data offer clear evidence of son-biased fertility stopping behavior. In 61% of all households in our sample the youngest child is a son. Moreover, consistent with son-biased fertility stopping behavior, we find that parents whose first child is a girl have on average 0.24 more children than parents whose first child is a boy[23].

We can then ask whether parents who have more boys than girls or parents whose youngest child is a son are also likely to have high levels of measured son bias. Reassuringly, this is indeed the case. The binned scatter plots in figure 5 show the individual level association of measured son bias and observed fertility outcomes. We find that participants with high levels of implicit and explicit son bias also have comparatively more son-biased fertility outcomes[24]. These results clearly shows that both our measures of son bias are systematically related to fertility outcomes.

We also find a significant association between measured son bias and the gender of a participant's first child, which means that interpreting the individual-level correlations shown in figure 5 requires some caution. Interestingly, a first-born daughter is associated with a reduction in implicit and explicit son bias by 0.12 and 0.28 standard deviations, respectively. Because sex-selective abortions at first parity are rare, we interpret this statistical association as a causal effect, suggesting that the experience of having daughters makes parents less son-biased [Blau et al., 2020, Dahl and Moretti, 2008, Lundberg, 2005][25]. Hence, the correlations in figure 5 are probably at least partly driven by a causal effect of

[22] While son-biased fertility stopping behavior cannot affect the aggregate sex ratio, it does increase the expected share of boys within any given household. To see this, consider a household that continues to have children until a son is born and then stops. If the first child is a son, the share of boys in this households will be 1. This occurs with probability 0.5 if we assume for simplicity that the natural sex ratio at birth is exactly 1. Similarly, for the share of boys in this household to be 0.5, the first child needs to be a girl and the second child a boy, which occurs with probability 0.5^2, and so on. The expected share of boys in this household is therefore given by $\sum_{x=1}^{\infty} 0.5^x \frac{1}{x} = \ln(2) \approx 0.693$.

[23] Some of these households may not yet have reached their desired family size. If we restrict our attention to households in which both parents agree that they do not want any more children, the share of households where the last child is a boy increases to 67% and the difference in average family sizes increases to 0.36 children.

[24] If we use D scores and index values in a horse-race regression, we find that both measures of son bias significantly predict fertility outcomes. The effect size is larger, however, for our index of explicit son bias. A possible explanation for the differential effect size is that fertility outcomes affect our measure of explicit son bias. If participants consider the gender of their youngest child when deciding on the optimal family composition of this child's future family, for example, the correlation between the index and the gender of the youngest child may be inflated.

[25] We estimate the difference in son bias between parents whose first child is a son and parents whose first child is a daughter by regressing our standardized measures of son bias on subject-type fixed effects and a dummy variable indicating whether the first child is a daughter. We also investigated whether the causal effect of having a first-born daughter differs for mothers and fathers, and we found no evidence for heterogeneous effects.

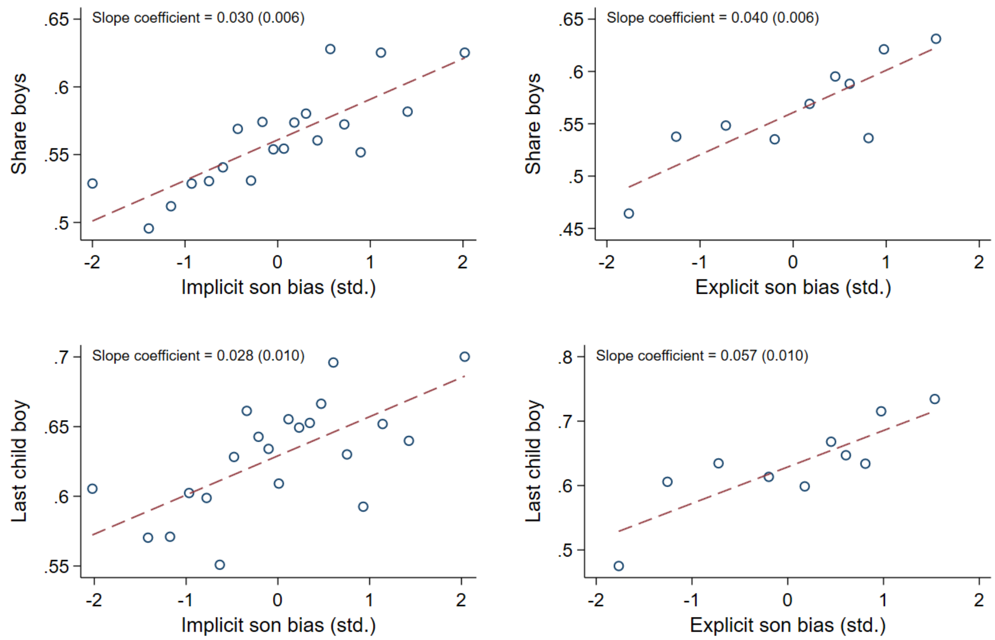

Figure 5: Binned scatter plot showing the individual-level association of measured son bias and fertility outcomes. Because only children are extremely rare in Armenia, we restrict our analysis to parents with two or more children. In all regression, we control for subject-type fixed effects. To facilitate comparisons of the effect sizes, we standardize our measures of implicit son bias (D scores) and explicit son bias (index values) to have zero mean and unit variance.

fertility outcomes on measured son bias.

The obvious way to conclusively demonstrate that our measures of son bias predict fertility outcomes would involve collecting data from prospective parents and then wait to record their fertility outcomes several years later. We have collected data from parents who already have children and therefore cannot offer this empirical test. Instead, we address the problem of reverse causality by making use of regional variation in son bias. The principal idea is that we can study the association of son bias and fertility outcomes at the community level while accounting for the effect of fertility outcomes on measured son bias at the household level[26]. We perform our analysis in two steps. First, we compute measures of son bias at the community level by regressing D scores and index values on community fixed effects while accounting for a full set of subject-type and family-composition fixed effects. Second, we use census data provided by the *Statistical Committee of the Republic of Armenia* to correlate the estimated community fixed effects with aggregate sex ratios in the 45 communities that were included in our study. Controlling for subject type in the first step addresses the problem that the distribution of subject types is not balanced across communities in our sample. This is important

[26]Finding a correlation between son bias and sex ratios measured at the community level requires that at least some of the heterogeneity in son bias is indeed across rather than within communities. As we will document in section IV, there is significant variation in son bias across regions. Moreover, geographic differences in sex ratios appear to be quite stable over time; Gegharkunik and Syunik were the regions with the highest and the lowest sex ratio among children under 16 in the 2011 census, and the same two regions are still at the extreme ends of the regional variation in the sex ratio at birth today [The Demographic Handbook of Armenia, 2020].

	Community-level sex ratio	
Correlations	Unweighted	Weighted
Implicit son bias	0.3091	0.3370
	(.0388)	(0.0236)
Explicit son bias	0.3931	0.3916
	(0.0076)	(0.0079)
Observations	45	45

p-values in parentheses.

Table 1: Community-level correlation of measured son bias and sex ratios at birth.

because male participants in our sample have significantly higher D scores than female participants, and we therefore need to account for the fact that the share of households in which we were able to interview husbands differs across communities[27]. We include family-composition fixed effects to absorb any effect of fertility outcomes on measured son bias at the household-level.

Sex ratios at the community level cannot be affected by son-biased fertility stopping behavior at the household level and therefore reflect the prevalence of sex-selective abortions. Table 1 shows that both our implicit as well as our explicit measures of son bias predict community-level sex ratios constructed from census data. The correlation coefficients are significant at standard significance levels and have the expected sign. Because the number of observations varies considerably across communities, we also report weighted correlation coefficients where the weights correspond to the number of observations in each community. The weighted and unweighted correlations are very similar. We interpret these correlations as evidence that our measures of explicit and our implicit son bias capture behaviorally relevant aspects of son bias, including a respondent's willingness to engage in sex-selection in favor of boys.

IV The structure of son bias in Armenia

Having established the internal and external validity of our measures of son bias, we can now use our data to investigate the structure of son bias in Armenia. We focus on important questions that require measures of son bias at the level of the individual decision-maker and cannot be answered using behavioral measures like sex ratios or parity progression ratios, which can only be estimated at the group level. We first explore the correlates of son bias, with a particular focus on how son bias differs between men and women and across different regions of Armenia. Given the documented gender gap in son bias, we then ask how common it is for husbands to be more son-biased than their wives. This second question is of importance for a policy-maker who considers aiming an intervention at men in an attempt to target the more son-biased parent in any given household. Our measures of son bias, which are relatively resistant to rationalization and social desirability biases, are well suited to answer these questions.

[27]See appendix A.1 for a detailed description of the sampling protocol.

IV.1 Gender differences, regional cultures, and socio-economic correlates of son bias

Does average son bias differ between men and women, or across different regions of Armenia? Both hypotheses appear natural. Different socialization processes for boys and girls and entrenched gender norms can instill differences in average son bias between men and women. Similarly, son bias may be rooted in local cultures that differ across regions. In fact, our implementation partners in Armenia often described son bias as a highly regional phenomenon, and census data confirm that sex ratios at birth differ between the three regions in our study. However, differences in sex ratios can also stem from differential access to sex-selection technology, and it is not clear ex ante whether significant differences in the degree of son bias exist between participants who live in Gegharkunik, a region with an extremely distorted sex ratio at birth, and participants living in Syunik, where the distortion of the sex ratio is less severe.

Below, we will show that regional differences in son bias exist, and they align with geographic variation in the sex ratio at birth. This finding, however, still leaves the question of whether differences in average son bias between Gegharkunik and Syunik really stem from differences in local cultures and norms. Alternatively, regional differences in son bias may simply reflect the socio-economic composition of the population in these two regions with residents of Gegharkunik being poorer and less educated on average than residents of Syunik. To sort out this question, we control for the independent effect of key socio-economic variables and study whether son bias in Armenia varies regionally in ways that are not explained by differences in the composition of the population[28].

Table 2 reports estimation results from ordinary least squares regressions that regress different measures of son bias on three sets of explanatory variables. The first set consists of a single dummy variable indicating whether the participant is male. The second set consists of two dummy variables indicating whether the participant lives in Gegharkunik or Syunik, the regions with the highest and the lowest sex ratios at birth in our sample. The omitted category is the capital region, Yerevan. In the third set, we include variables describing a participant's socio-economic status. This set includes measures of household income, education level, and age, as well as dummies for a participant's profession[29]. Besides reporting the coefficients for each of these regressors, we also report Owen values from a Shapley decomposition of the coefficient of determination into the respective components explained by each set of regressors [Huettner et al., 2012, Shorrocks, 2013][30].

Using the D scores from our implicit association test as the dependent variable, we find that male participants are much more son-biased than female participants. Being male increases the D scores by about half a standard deviation. Furthermore, participants living in Gegharkunik, the region with the highest sex ratio at birth in our sample, are significantly more son-biased than participants living in Syunik, the region with the lowest sex ratio at birth. This regional gap is not explained by differences in income, education, age, or profession and appears consistent with a local culture of strong son bias in Gegharkunik. However, it is important to note that the regional differences in son bias are small compared to the gender difference, which is about three times as large. Interestingly, none of the

[28] Because Armenia is very homogeneous ethnically, we ignore difference in fertility behavior across ethnic groups and focus on the composition of the population in terms of socio-economic characteristics.

[29] See appendix A.6 for a description of how we construct the education and income variable.

[30] Differences in overall R^2 between the specifications likely reflect differences in the relative amount of noise in the dependent variable. Implicit association tests, for example, are designed to be immune to social desirability bias, but likely constitute a relatively noisy measure of latent son bias. This does not, however, affect the interpretation of the Shapley values, which are unaffected by classical measurement error in the dependent variable.

Dep. Variable	*D* scores (std.)		Index (std.)		Personal view		Perceived norm	
	OLS	*Shapley*	OLS	*Shapley*	OLS	*Shapley*	OLS	*Shapley*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Respondent type		*74.9%*		*20.3%*		*0.6%*		*0.2%*
Male	0.527***		0.439***		−0.025		−0.010	
	(0.057)		(0.053)		(0.064)		(0.053)	
Region		*9.2%*		*57.8%*		*46.4%*		*92.9%*
Gegharkunik	0.176***		0.492***		0.386***		0.086	
	(0.065)		(0.060)		(0.073)		(0.061)	
Syunik	−0.040		−0.236***		−0.334***		−0.558***	
	(0.059)		(0.055)		(0.067)		(0.055)	
Socio-economic status		*16.0%*		*22.0%*		*53.0%*		*6.9%*
Income Level	0.012		0.076***		−0.075***		0.006	
	(0.021)		(0.020)		(0.024)		(0.020)	
Education Level	−0.031		−0.075***		−0.149***		−0.019	
	(0.029)		(0.027)		0.033)		0.027)	
Year of birth	−0.003		−0.004*		−0.015***		−0.002	
	(0.002)		(0.002)		(0.002)		(0.002)	
Agriculture	0.010		−0.127		0.040		−0.015	
	(0.095)		(0.088)		(0.107)		(0.088)	
Commerce and private business	−0.060		−0.110		0.176*		0.072	
	(0.084)		(0.078)		(0.094)		(0.078)	
Construction	0.026		−0.049		0.364**		0.109	
	(0.132)		(0.123)		(0.149)		(0.123)	
Government, IGO, or NGO employee	0.034		−0.126**		0.113		0.087	
	(0.065)		(0.061)		(0.074)		(0.061)	
Pay by the day	0.048		−0.226*		−0.259*		0.050	
	(0.131)		(0.122)		(0.148)		(0.122)	
Constant	5.037		7.187*		29.981***		4.174	
	(4.015)		3.733)		4.532)		3.743)	
N	1,804		1,804		1,804		1,804	
Overall R^2	0.065		0.146		0.130		0.087	
F-stat model	11.363		27.904		24.405		15.534	

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Regression results and Shapley decomposition. The results in this table are based on our *representative sample*, which excludes households living in communities that we have specifically targeted because of their extreme sex ratios at birth. See appendix A.1 for details about our two-stage sampling approach. *D* scores and index values are standardized to have unit variance, so that coefficients on correlates are comparable. The variables in columns (5) through (8) are constructed from responses to the questions about the Armenian saying discussed in the main text. We encode the responses by assigning values between 1 (strongly disagree) and 5 (strongly agree). For occupations, the excluded category is "House work or other".

socio-economic variables are significant at standard levels. Perhaps even more strikingly, the Shapley decomposition in column (2) reveals that the gender of the participant alone accounts for almost three fourths of the explained variation in son bias, while regions account for less than 10%, and our entire set of socio-economic variables accounts for only 16%.

How does our explicit measure of son bias compare to this? In columns (3) and (4) we show the same regression but replace the D scores with our index values. Consistent with our previous results, we again find that male participants are significantly more son-biased than female participant, and inhabitants of Gegharkunik are more son-biased than inhabitants of Syunik. The gender gaps in explicit and implicit son bias are of comparable magnitude, while regional differences are more pronounced for our measure of explicit son bias. Socio-economic variables do play some role in explaining son-biased fertility preferences, with more educated participants and employees of the government or NGOs being less son-biased, and participants with higher household income being more son-biased. However, we again find that socio-economic variables do not explain much of the variation in son bias, while region and gender alone account for almost 80% of the explained variation[31].

How should we interpret the fact that regional differences appear more pronounced when we study the distribution of our measure of explicit son bias compared to the results we obtain based on our implicit association tests? One interpretation is that explicit fertility preferences, based on which we construct the index, are subject to social expectation while D scores are not. As explained in section III.2, our measure of explicit son bias is constructed based on self-reported gender preferences for the future family of the youngest child in the household. If participants live in an environment where son bias is the norm, they may want to report more son-biased gender preferences – either because social desirability induces them to do so, or because they believe that having sons will allow their offspring to thrive in a son-biased environment. To understand if this is a plausible interpretation, we study regional variation son-biased norms.

It is often argued that preferences for male children in patrilineal societies like Armenia stem from the notion that sons ensure the continuation of the family lineage, while daughters leave upon marriage to live with the families of their spouses (see for example Murphy et al. [2011]). To understand whether the importance of patrilineal norms differs between the three regions in our sample, we asked participants to what extent they agree with the Armenian saying *"the son is ours, the daughter belongs to the others."* We also asked our participants to what extent they think that other parents in Armenia agree with this saying, which allows us to study the perception of norms [Bicchieri, 2005]. The order in which the two questions were presented was randomized across participants, and we scored the answers on a five-level Likert item ranging from strongly disagree (1) to strongly agree (5). 57% of respondents think that other parents in Armenia either agree or strongly agree, but only 36% of respondents in our sample actually do so themselves.

In column (5) and (6), we repeat the regression analysis using as the dependent variable the degree of personal agreement with the Armenian saying. We find that people in Gegharkunik indeed agree more strongly with the saying compared to people in Syunik. Socio-economic variables also play a role with richer, more educated, and younger people agreeing less. How a participants answers

[31] Finding little correlation between son bias and socio-economic variables contrasts with studies in other countries that find significant associations between the sex ratio at birth and parental education, profession, or wealth [Siddhanta S. et al., 2003, Bhat and Zavier, 2007, Chun et al., 2009, Guilmoto and Ren, 2011]. It is important to remember, however, that socio-economic variables can predict the sex ratio at birth even in contexts where son bias does not vary across socio-economic groups. Richer parents may be more likely to have sons, for example, if insufficient financial means make it harder to engage in sex selection.

questions about her preferred family composition may also depend on the perceived importance of son-biased norms. In columns (7) and (8), we therefore make use of the second question where we ask participants how strongly they think that other parents agree with the saying. It turns out that whether a participants thinks that other parents agree with the saying depends almost entirely on the region that the participant lives in, with participants in Syunik thinking much less frequently that other parents would agree with the saying relative to participants in Gegharkunik and in Yerevan. To the extent that responses to our survey questions are affected by social expectations, these differences in perceived norms suggest that our measure of explicit son bias likely overstates regional differences.

To summarize, using D scores as our preferred measure of son bias we find that the greatest differences in son bias occur between male and female participants. While regional differences in son bias do exist, they are small compared to the gender difference. Socio-economic variables on the other hand cannot explain differences in son bias.

IV.2 Spousal differences in son bias

In this last section we discuss the policy relevance of the documented gender gap in son bias. We consider the case of a policy-maker who has access to an intervention that is believed to greatly reduce or eliminate son bias among individuals exposed to the intervention[32]. The policy-maker has limited resources and, knowing that son bias varies across individuals, must decide what segment of the population to target in an effort to maximize the impact of the intervention. Any targeting strategy that exposes individuals with little or no son bias to the intervention is inefficient. Unfortunately, direct measures of fertility preferences are typically not available, and the fact that son bias appears largely unrelated to socio-economic variables limits the scope for identifying relatively more son-biased households. The significant gender gap in son bias documented in the previous section, however, raises the prospect that by targeting men, the intervention can be focused on the more son-biased parent in any given household. The strategy of targeting the more son-biased parent in any given household promises to be particularly effective if the intervention is expected to produce local spillovers inside the household [Efferson et al., 2020][33].

To assess the extent to which a policy maker can rely on gender as a proxy for relative son bias, we would like to know the probability that a randomly sampled husband is more son-biased than his wife. This probability depends on the difference in average son bias between men and women, the variation in son bias among men and women, and the degree to which the fertility preferences of husbands and wives are correlated. To compute this probability we leverage the fact that, whenever possible, we have collected data from both spouses[34]. We begin our analysis with the rough-and-ready approach

[32]For example, the intervention could be an information campaign, a focus group workshop, or a conditional cash transfer. Alternatively, the policy-maker may try to affect the sex ratio at birth through a supply side intervention that limits access to sex-selection technology. This paper, however, focuses on the demand for sons, and we therefore consider the case of a policy-maker committed to reducing son bias.

[33]This conclusion rests on the assumption that the intervention is equally effective for all son-biased individuals. If, in contrast, the effects of the intervention are heterogeneous, the policy maker must manage this additional source of complexity, but a reliable measure of or proxy for individual preferences remains extremely useful. Whether the policy-maker can lower the sex ratio at birth by targeting men also depends on the question of whether fathers or mothers exert more control over decisions regarding family planning. In appendix A.7, we show that fertility outcomes at the household level are correlated with measured son bias of both mothers and fathers, suggesting that the fertility preferences of both parents matter.

[34]For the analyses in this section, we drop the data from husbands' mothers. Similarly, we also drop all data from households, in which we were able to collect data only from one of the spouses.

of simply counting how often the level of measured son bias for a male participant exceeds that of his wife. Interestingly, despite the large gender difference in average son bias documented in the previous section, D scores of husbands, our measure of implicit son bias, exceed those of their wives only 69 out of 100 times. Our measure of explicit son bias finds husbands more son-biased than their wives only 54 out of 100 times. One possible interpretation of this result is that heterogeneity in son bias among individuals of a given sex overshadows the documented difference in average son bias between men and women. Unless the degree of spousal correlation is very high, this would imply that husbands are in many cases actually less son-biased than their wives.

Another interpretation, however, is that unlike height or weight, which can be measured precisely, most psychological traits like tastes or attitudes can only be measured with some degree of noise. For example, despite the obvious virtues of implicit association tests, it is a challenge for implicit measures to achieve high levels of test-retest reliability [Cunningham et al., 2001, Nosek et al., 2007]. Similarly, not all the variation in our index measure will reflect true variation in latent son bias. While random noise does not affect our estimates of the gender gap in son bias, it is important to note that imperfect reliability biases the degree of spousal correlation in measured son bias towards zero and implies that a simple comparison of measured son bias among husbands and wives understates the reliability of gender as a predictor of relative son bias.

A first step in dealing with noise is to combine D scores and index values in a composite measure of son bias that has higher reliability. We construct this composite measure of son bias by taking the average of an individual's D score and index value, after having normalized both measures to have unit variance. As expected, we find that this step increases the share of households, in which measured son bias is larger for the husband relative to his wife. The degree of spousal correlation in our composite measure of son bias is 0.17, after accounting fo the gender composition of children in the household, and in 72% of households the degree of measured son bias is larger for the husband compared to his wife. However, if there is noise in our D scores and index values, then this composite measure will also fall short of being a perfectly reliable measure of latent son bias. As a consequence, we still underestimate the degree to which latent son bias is correlated among spouses, and we overestimate the share of households, in which the wife is more son-biased than her husband. In a second step, we therefore deal with noise in our composite measure of son bias more formally.

Let us postulate that the relationship between measured son bias and latent son bias can be written as

$$measuredBias_i = latentBias_i + \varepsilon_i$$

where ε_i is an independent error term that captures noise. We assume that latent son bias of husband and wife is jointly normally distributed with mean, $\left(\begin{smallmatrix}\mu_{husband} \\ \mu_{wife}\end{smallmatrix}\right)$, and covariance matrix, $\left(\begin{smallmatrix}\sigma^2_{husband} & \sigma \\ \sigma & \sigma^2_{wife}\end{smallmatrix}\right)$. The probability that a husband's degree of latent son bias exceeds that of his wife is then given by

$$F\left(\frac{\mu_{husband} - \mu_{wife}}{\sqrt{\sigma^2_{husband} + \sigma^2_{wife} - 2\rho\sigma_{husband}\sigma_{wife}}}\right)$$

where $F(\cdot)$ is the cumulative distribution function of a standard normal distribution and ρ is the degree of spousal correlation in latent son bias. The expected levels of son bias among husbands and wives, which appear in the numerator, can be estimated directly from the data. Moreover, for a given assumption on the variance of ε, we can also use the data to derive estimates for $\sigma_{husband}$, σ_{wife}, and

	Reliability coefficient		
	0.75	0.5	0.25
Spousal correlation, ρ	0.20	0.30	0.60
Prob($latentBias^{husband} > latentBias^{wife}$)	0.75	0.81	0.95

Table 3: Spousal correlation in latent son bias and the probability that a randomly sampled husband is more son-biased that his wife.

ρ[35].

We consider three alternative scenarios for the relative amount of noise in our composite measure of son bias. In a low noise scenario, we assume that noise accounts for a quarter of the variation in measured son bias among men and women, respectively. In a medium noise scenario, we assume that half of the variation in measured son bias is due to noise. Finally, in a high noise scenario, we assume that noise accounts for three quarters of the variation in measured son bias. These scenarios correspond to assuming that the reliability coefficient of our composite measure, defined as the ratio of variation in latent son bias to total variation in measured son bias, takes a value in $\{0.75, 0.5, 0.25\}$.

Table 3 reports the degree of spousal correlation implied by a given assumption on the relative importance of noise, as well as the associated probability that a husband is more son-biased that his wife. Moving from the low noise to the high noise scenario, the implied degree of spousal correlation in latent son bias increases by a factor of three. Similarly, the implied percentage of households, in which husbands are more son-biased than their wives increases from 75% to 95%. Hence, gender is a fairly reliable proxy for relative son bias among parents, especially if one believes that latent son bias is measured with a significant degree of noise. For the policy-maker who tries to make efficient use of her resources by always targeting the more son-biased parent in any given household, these results suggest that focusing on men can be a simple strategy to approximate this goal.

V Discussion

In this paper we implement and validate novel measures of implicit and explicit son bias, and we demonstrate their usefulness in revealing the structure of son bias in Armenia. Our novel implicit association tests are designed to reveal son bias while avoiding the problems associated with directly interrogating individuals about their preferred fertility outcomes. We complement this measure of implicit son bias with a measure of explicit son bias based on questionnaire items, which builds on recent work by Jayachandran [2017] and overcomes important biases characterizing past research. We demonstrate that our measures of implicit and explicit son bias are internally consistent, and that they correlate with son-biased fertility outcomes at both the household and the community level.

An important strength of our research design is the fact that we measure fertility preferences at the level of individual decision-makers. By extension, this opens a path to analyzing how individual preferences vary within households, across regions, and with socioeconomic variables. Because we

[35] The standard deviation of latent son bias among husbands and wives can be computed as $\sigma_k = \sqrt{\mathrm{Var}(measuredBias^k) - \mathrm{Var}(\varepsilon)}$ where $k \in \{husband, wife\}$. The spousal correlation in latent son bias can be estimated as $\rho = \left(\sigma_{husband}\sigma_{wife}\right)^{-1}\tilde{\sigma}$ where $\tilde{\sigma}$ denotes the spousal covariance in measured son bias conditional on the gender composition of children in the household.

measure fertility preferences of both parents, we can study gender differences in son bias. Our analyses reveal that men are on average significantly more son-biased than women. This finding could not have been obtained from analyses that focus exclusively on fertility outcomes, as these are realized at the household level and do not allow the researcher to distinguish between the preferences of fathers and mothers. Our finding of an important gender gap in son bias echo results in other studies, which show that fathers tend to spend more time with their sons than daughter and that sons increase marital stability. Interpreting gender differences in the behavior of parents is difficult, however, because fathers may have a comparative advantage in raising sons, or parents may believe that male role models are more important for boys than girls [Lundberg, 2005, Dahl and Moretti, 2008, Blau et al., 2020]. In contrast to studies that infer son bias from behavior, our methodology allows us to measure spousal differences in son bias directly.

Our analyses also reveal regional differences in son bias that align with geographic variation in the sex ratio at birth. These regional differences persist after accounting for the composition of the population in terms of age, income, and education. Investigating what explains regional difference in son bias promises to be an insightful avenue for future research. One possible explanation is migration. A significant fraction of men in Armenia spends part of the year abroad, most often in Russia, in search of better job opportunities. The resulting absence of men and the economic dependency on remittances may well affect son bias. Gegharkunik, the region with the most distorted sex ratio at birth in our sample, has seen particularly high levels of out-migration in the early 2000s, and migration still plays an important role today [Yeganyan and Shahnazaryan, 2004, Sevoyan and Agadjanian, 2010, Guilmoto, 2013, Agadjanian and Sevoyan, 2014]. This regional variation in migration patterns is reflected in our data. Among sampled households in Yerevan and Syunik only 6% have a father who regularly works abroad, while this is the case for 32% of all sampled households in Gegharkunik. These number are only correlational, of course, but they raise questions about the causal impact of migration on fertility preferences.

Inter-regional migration can also provide important insights about the nature of son bias. For example, the question of whether son bias is transmitted vertically within families or horizontally under the influence of social norms could be studied by comparing families that have lived in the same region for several generation to families that have only recently immigrated from another region with either lower or higher average levels of son bias. If vertical transmission is more important than horizontal transmission, then son bias among the descendants of migrants should remain close to average son bias in the region of origin.

Despite our extensive efforts to provide privacy for our participants and to minimize the risk of social desirability bias, we have clear indications that sex-selective abortions are under-reported. This serves as an important reminder that social desirability bias is a serious concern for research on sex-biased fertility preferences. It also underscores the value of our implicit association tests in establishing the robustness of our main results. By showing that implicit association tests can be used to reveal son bias, our paper contributes to the growing literature that uses implicit association tests in a field setting to measure attitudes about sensitive topics in which respondents may not be prepared to explicitly reveal their attitudes [Beaman et al., 2009, Lowes et al., 2015, Efferson et al., 2015, Vogt et al., 2016].

Our findings of a significant gender gap in son bias can inform the design of effective policy interventions. In most situations, the policy-maker will not have access to individual-level data on fertility preferences, and we discuss the reliability of gender as a proxy for son bias. We estimate the degree of spousal correlation in son bias and argue that most husbands in our sample are more son-biased than their wives. This provides a handy proxy for the policy-maker who tries to make efficient use of her resources by always targeting the more son-biased parent in any given household. Our results suggest that targeting husbands is a ready shortcut that is likely to accomplish exactly that.

References

V. Agadjanian and A. Sevoyan. Embedding or uprooting? the effects of international labour migration on rural households in armenia. *International Migration*, 52(5):29–46, 2014.

O. Altindag. Son preference, fertility decline, and the nonmissing girls of turkey. *Demography*, 53(2): 541–566, 2016.

J. B. Asendorpf, R. Banse, and D. Mücke. Double dissociation between implicit and explicit personality self-concept: the case of shy behavior. *Journal of Personality and Social Psychology*, 83(2):380–393, Aug. 2002.

D. Basu and R. De Jong. Son targeting fertility behavior: some consequences and determinants. *Demography*, 47(2):521–536, 2010.

L. Beaman, R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova. Powerful women: does exposure reduce bias? *The Quarterly journal of economics*, 124(4):1497–1540, 2009.

P. M. Bhat and A. F. Zavier. Factors influencing the use of prenatal diagnostic techniques and the sex ratio at birth in india. *Economic and Political Weekly*, pages 2292–2303, 2007.

C. Bicchieri. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press, 2005.

F. D. Blau, L. M. Kahn, P. Brummund, J. Cook, and M. Larson-Koester. Is there still son preference in the united states? *Journal of Population Economics*, 33(3):709–750, 2020.

J. Bongaarts. The measurement of wanted fertility. *Population and development review*, 16(3):487–506, 1990.

M. D. Cattaneo, R. K. Crump, M. H. Farrell, and Y. Feng. On binscatter. Feb. 2019.

F. Chao, P. Gerland, A. R. Cook, and L. Alkema. Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels. *Proceedings of the National Academy of Sciences*, 116(19):9303–9311, 2019.

H.-R. Chun, I.-H. Kim, and Y.-H. Khang. Trends in sex ratio at birth according to parental social positions: results from vital statistics birth, 1981-2004 in korea. *Journal of Preventive Medicine and Public Health*, 42(2):143–150, 2009.

S. Clark. Son preference and sex composition of children: Evidence from india. *Demography*, 37(1): 95–108, 2000.

L. C. Coombs. The measurement of family size preferences and subsequent fertility. *Demography*, 11 (4):587–611, 1974.

W. A. Cunningham, K. J. Preacher, and M. R. Banaji. Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological science*, 12(2):163–170, 2001.

G. B. Dahl and E. Moretti. The demand for sons. *The Review of Economic Studies*, 75(4):1085–1120, 2008.

H. P. David. Abortion in europe, 1920-91: a public health perspective. *Studies in family planning*, 23 (1):1–22, Jan. 1992.

G. Duthé, F. Meslé, J. Vallin, I. Badurashvili, and K. Kuyumjyan. High sex ratios at birth in the caucasus: Modern technology to satisfy old desires. *Population and Development Review*, 38(3): 487–501, Sept. 2012.

C. Efferson, S. Vogt, A. Elhadi, H. E. F. Ahmed, and E. Fehr. Female genital cutting is not a social coordination norm. *Science*, 349(6255):1446–1447, 2015.

C. Efferson, M. Schief, and S. Vogt. Preferences for sons in armenia, Jul 2018. URL `https://osf.io/pbfej/`.

C. Efferson, S. Vogt, and E. Fehr. The promise and the peril of using social influence to reverse harmful traditions. *Nature human behaviour*, 4(1):55–68, 2020.

B. Egloff and S. C. Schmukle. Predictive validity of an implicit association test for assessing anxiety. *Journal of Personality and Social Psychology*, 83(6):1441–1455, Dec. 2002.

K. A. Fenton, A. M. Johnson, S. McManus, and B. Erens. Measuring sexual behaviour: methodological challenges in survey research. *Sexually transmitted infections*, 77(2):84–92, 2001.

D. Filmer, J. Friedman, and N. Schady. Development, modernization, and childbearing: The role of family sex composition. *The World Bank Economic Review*, 23(3):371–398, 2009.

K. Fuse. Variations in attitudinal gender preferences for children across 50 less-developed countries. *Demographic Research*, 23:1031–1048, 2010.

A. G. Greenwald and B. A. Nosek. Health of the implicit association test at age 3. 2001.

A. G. Greenwald, D. E. McGhee, and J. L. Schwartz. Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6):1464–1480, June 1998.

A. G. Greenwald, B. A. Nosek, and M. R. Banaji. Understanding and using the implicit association test: I. an improved scoring algorithm. *Journal of personality and social psychology*, 85(2):197, 2003.

A. G. Greenwald, T. A. Poehlman, E. L. Uhlmann, and M. R. Banaji. Understanding and using the implicit association test: III. meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1):17–41, July 2009.

C. Guilmoto. Sex imbalances at birth in armenia: Demographic evidence and analysis, 2013.

C. Z. Guilmoto. Gender bias in reproductive behaviour in georgia, indonesia, and vietnam: An application of the own-children method. *Population studies*, 71(3):265–279, 2017.

C. Z. Guilmoto and Q. Ren. Socio-economic differentials in birth masculinity in china. *Development and change*, 42(5):1269–1296, 2011.

C. Z. Guilmoto and J. Tove. The masculinization of births. overview and current knowledge. *Population*, 70(2):185–243, 2015.

A. Hassan, S. B. Schaffnit, R. Sear, M. Urassa, and D. W. Lawson. Fathers favour sons, mothers don't discriminate: Sex-biased parental care in northwestern tanzania. *Evolutionary Human Sciences*, 1, 2019.

S. A. Hohmann, C. A. Lefèvre, and M. L. Garenne. A framework for analyzing sex-selective abortion: the example of changing sex ratios in southern caucasus. *International journal of women's health*, 6:889, 2014.

F. Huettner, M. Sunder, et al. Axiomatic arguments for decomposing goodness of fit according to shapley and owen values. *Electronic Journal of Statistics*, 6:1239–1250, 2012.

International Center for Human Development. Monitoring public policy and programs to prevent gender-biased sex selection and sex-selective abortions in the republic of armenia. `https://ichd.org/download.php?f=870&fc=Monitoring%20Report%20(Eng%20-%20web%20version).pdf`, 2016.

S. Jayachandran. Fertility decline and missing women. *American Economic Journal: Applied Economics*, 9(1):118–139, 2017.

A. Jilozian and V. Agadjanian. Is induced abortion really declining in armenia? *Stud. Fam. Plann.*, 47(2):163–178, June 2016.

A. Karpinski and J. L. Hilton. Attitudes and the implicit association test. *Journal of Personality and Social Psychology*, 81(5):774–788, Nov. 2001.

R. Kashyap and F. Villavicencio. The dynamics of son preference, technology diffusion, and fertility decline underlying distorted sex ratios at birth: A simulation approach. *Demography*, 53(5):1261–1281, 2016.

L. Khachatryan. Survey on identification of skills needed in the agrarian sector and motivation of youth to pursue a career in the agricultural sector in armenia. 2019.

D.-Y. Kim. Voluntary controllability of the implicit association test (IAT). *Social Psychology Quarterly*, 66(1):83–96, 2003.

S. Lowes, N. Nunn, J. A. Robinson, and J. Weigel. Understanding ethnic identity in africa: Evidence from the implicit association test (iat). *American Economic Review*, 105(5):340–45, 2015.

S. Lundberg. Sons, daughters, and parental behaviour. *Oxford Review of Economic Policy*, 21(3): 340–356, 2005.

V. Mazza, C. Falcinelli, S. Paganelli, G. Contu, S. Mantuano, S. Battafarano, A. Forabosco, and A. Volpe. Sonographic early fetal gender assignment: a longitudinal study in pregnancies after in vitro fertilization. *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 17(6):513–516, 2001.

G. H. McClelland. Determining the impact of sex preferences on fertility: a consideration of parity progression ratio, dominance, and stopping rule measures. *Demography*, 16(3):377–388, 1979.

G. H. McClelland. Measuring sex preferences and their effects on fertility. In N. G. Bennett, editor, *Sex Selection of Children*, pages 13–46. Academic Press, 1983.

S. G. McFarland and Z. Crouch. A cognitive skill confound on the implicit association test. *Social Cognition*, 20(6):483–510, 2002.

F. Meslé, J. Vallin, and I. Badurashvili. A sharp increase in sex ratio at birth in the caucasus. why? how. *Watering the Neighbour's Garden: The Growing Demographic Female Deficit in Asia, Paris: Committee for International Cooperation in National Research in Demography*, pages 73–88, 2007.

M. Michael, L. King, L. Guo, M. McKee, E. Richardson, and D. Stuckler. The mystery of missing female children in the caucasus: an analysis of sex ratios by birth order. *International perspectives on sexual and reproductive health*, pages 97–102, 2013.

R. Murphy, R. Tao, and X. Lu. Son preference in rural china: Patrilineal families and socioeconomic change. *Population and development review*, 37(4):665–690, 2011.

J. Norling. Measuring heterogeneity in preferences over the sex of children. *Journal of development economics*, 135:199–221, 2018.

B. A. Nosek, A. G. Greenwald, and M. R. Banaji. Understanding and using the implicit association test: Ii. method variables and construct validity. *Personality and Social Psychology Bulletin*, 31(2): 166–180, 2005.

B. A. Nosek, A. G. Greenwald, and M. R. Banaji. The implicit association test at age 7: a methodological and conceptual review. In J. A. Bargh, editor, *Automatic Processes in Social Thinking and Behavior*, pages 265–292. Psychology Press, 2007.

L. I. Remennick. Epidemology and determinants of induced abortion in the USSR. *Social science & medicine*, 33(7):841–848, 1991.

M.-C. Robitaille. Determinants of stated son preference in india: are men and women different? *The Journal of Development Studies*, 49(5):657–669, 2013.

M.-C. Robitaille and I. Chatterjee. Sex-selective abortions and infant mortality in india: the role of parents' stated son preference. *The Journal of Development Studies*, 54(1):47–56, 2018.

M. R. Rosenzweig and K. I. Wolpin. Maternal expectations and ex post rationalizations: the usefulness of survey information on the wantedness of children. *Journal of human resources*, pages 205–229, 1993.

A. Sen. Missing women. *BMJ: British Medical Journal*, 304(6827):587, 1992.

A. Sevoyan and V. Agadjanian. Male migration, women left behind, and sexually transmitted diseases in armenia. *International Migration Review*, 44(2):354–375, 2010.

Sexuality Policy Watch. Armenia: Changes in the law on abortion, 2016. URL `https://sxpolitics.org/armenia-changes-in-the-law-on-abortion/16057`.

A. F. Shorrocks. Decomposition procedures for distributional analysis: a unified framework based on the shapley value. *Journal of Economic Inequality*, 11(1):99, 2013.

Siddhanta S., Nandy D., and Agnihotri S. B. Special statistics - 36: Sex ratios and 'prosperity effect'. what do NSSO data reveal? *Economic and Political Weekly*, 38(41):4381–4404, 2003.

The Demographic Handbook of Armenia. *Statistical Committee of the Republic of Armenia*, 2020.

United Nations Population Division. World population prospects 2019. 2019.

S. Vogt, N. A. M. Zaid, H. E. F. Ahmed, E. Fehr, and C. Efferson. Changing cultural attitudes towards female genital cutting. *Nature*, 538(7626):506–509, 2016.

C. F. Westoff and N. B. Ryder. The predictive validity of reproductive intentions. *Demography*, 14(4): 431–453, 1977.

K. Yamaguchi. A formal theory for male-preferring stopping rules of childbearing: Sex differences in birth order and in the number of siblings. *Demography*, 26(3):451–465, 1989.

R. Yeganyan and N. Shahnazaryan. Labor migration from armenia: a review of literature. *Organization for Security and Cooperation in Europe. International Organization for Migration. Yerevan.(In Russian)*, 2004.

SUPPLEMENTARY INFORMATION

A Appendix

A.1 Sampling Protocol and Data Collection

Target population The study focuses on three areas: Yerevan, Gegharkunik, and Syunik. Yerevan is the capital region, while Gegharkunik and Syunik are the regions with the highest and lowest sex ratio among children below the age of 16 in the 2011 national census. Within each region, we randomly sampled married couples with at least one child under 16 living at home according to the two-stage sampling approach described below. For a sampled couple, we collected data with both the husband, the wife, and the husband's mother during the same visit. If a couple was sampled, but either the wife or the husband had deceased or was living in a distant region or abroad for the duration of the study, we still collected data with the other spouse (and the husband's mother if available). In terms of sampling, we call such couples "incomplete couples." The rationale behind collecting data from incomplete coupes was to avoid selection bias that may have resulted if complete couples and incomplete couples were structurally different in terms of any variable of interest. If a couple was sampled, and both spouses were available at some time during the study, participation of both souses in the data collection was a pre-condition for participation of the given household in our study. We call such couples "complete couples." Finally, the husband's mother was included in the study whenever she was available on the day of data collection and she agreed to participate.

The study took place between May 2017 and March 2019. In each of the three regions, we planned to collect data from 900 participants leading to a total sample size of 2700 participants[36]. We sampled households based on a two-stage sampling approach that oversampled communities with especially high or low sex ratios at birth to maximize variation in the outcomes of interest, while at the same generating a subsample that is representative of the target population in the three regions. The details of the sampling protocol are described below. Our final sample consists of 1,212 households, in which we collected data from 2.23 household members on average. Overall, 44% of our respondents are wives, 33% are husbands, and 23% are mothers-in-law. In 73% of all households we collected data from both the husband and the wife and in 34% we have data from all three eligible household members.

Representativeness While our data come from an interesting set of regions made up by the capital as well as two region at the opposite extreme ends of the regional variation in the sex ratio at birth, no claim is made about the representativeness of our sample for the country of Armenia. We explore the extent to which households in our sample are nationally representative in terms of their propensity to engage in sex selection by comparing sex ratios at various parities in our data to national figures. In particular, we focus on one section in our questionnaire in which wives anonymously provided their complete reproductive histories and from this we calculate the sex ratios at birth that we then compare to those reported from a nationally representative sample in Guilmoto [2013]. Figure A.1 shows the results. Guilmoto [2013] reports sex ratios for births in the late 1990s and for births during the years 2001-2010. The latter category is closest to our sampling frame. The comparison provides a very close match for parities up to three and suggests that in terms of the prevalence of sex-selective abortions

[36] Due to a facilitator's mistake in entering the subject identification number we lost one observation, which leads to a final sample size of 2699 participants.

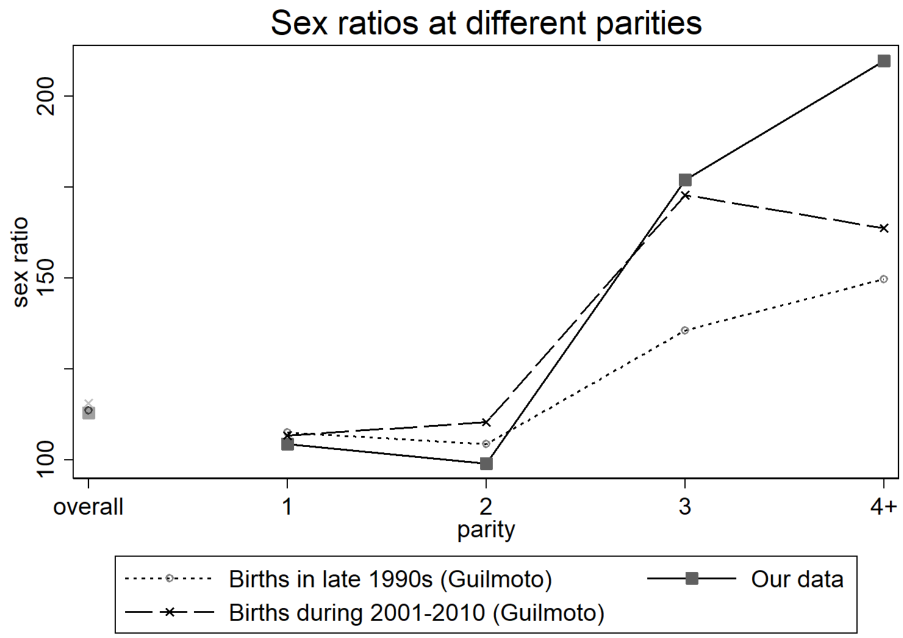

Figure A.1: Comparison of sex ratios in our data and sex ratios obtained from national birth records.

our sample may indeed be representative of Armenia at large.

Two-stage Sampling We implemented a two-stage sampling approach based on complete records of the sex of all children born in the three regions of interest. The data was provided by the *Statistical Committee of the Republic of Armenia* and is taken from the *2011 National Census*. In the first stage of our sampling strategy, we selected communities from within each region. In the second stage, we then sampled addresses from within a given sampled community. In each region, we created three distinct samples. First, we created a sample that is representative for our target population at the regional level. We will call this sample the *representative sample* and two third of our participants are part of this sample. Next, we specifically sampled addresses from "extreme" communities that are characterized by unusually high or unusually low observed sex ratios at birth. We will call these samples *lower tail sample* and *upper tail sample* respectively to indicate that for these sample we have specifically chosen communities from the tails of the distribution of community-level sex ratios at birth. Those samples comprise one sixth of our participants each.

Determining the Sample Size We committed to collecting data from 900 participants per region. Assuming that we would on average be able to collect data with 2.25 participants per household, we concluded that we would need to work in approximately 400 households in a given region. Moreover, we assumed that only every third address that we sampled would eventually lead to a successful recruitment of the household living at that address. Finally, we wanted to avoid creating too high a concentration of sampled household in any given location and we therefore restricted ourselves to sample only 1 out of every 5 addresses that met our sampling requirements. Given these restrictions, we concluded that our *representative sample* should consist of 800 households while the *lower tail* *sample* and *upper tail sample* should consist of 200 households each.

Constructing the Representative Sample We constructed the *representative sample* by consecutively picking communities from within a given Marz and summing up the number of eligible households in the sampled communities. Communities were picked at random and we stopped picking additional communities once we had reached the target of 800 sampled households. In the second stage we then randomly sampled 20% of the households in those communities.

Constructing the Upper Tail Sample The task is to identify the communities whose inhabitants are most prone to selectively aborting female fetuses. However, we do not observe sex-selective abortions directly and must rely instead on data about sex ratios at birth, which are imperfect proxies for the prevalence of sex-selective abortions. Moreover, we did not want to simply select the communities with the highest sex ratios, because for many communities we observe only relatively few births and distorted sex ratios may then be the outcome of chance rather than a true reflection of widespread son bias. Hence, our aim was to identify the communities whose sex ratios at birth suggest the highest prevalence of sex selection provided that there is adequate evidence that the observed sex ratios truly reflect sex-selective abortions.

We think of the observed sex ratio in a given community as one realization of a data generating process. The data generating process specifies the probability that a child is born as a boy and therefore captures the underlying prevalence of sex selection. Let p denote this probability, which is assumed to vary across communities. With this interpretation at hand, we can formulate a null hypothesis specifying the probability that a child is born as a boy in a given community and we can perform a one sided binomial test to decide whether the null hypothesis can be rejected in favor of the alternative hypothesis that assumes the data generating process assigns a higher probability to the birth of a boy.

To select communities for the *upper tail sample*, we first excluded all the communities that had already been sampled as part of the *representative sample* and the searched for the communities for which we could reject the most 'extreme' null hypothesis. More precisely, we set the required significance level to 10% (i.e. $\alpha = 0.1$) and looked for the highest probability p^{null} subject to the constraint that the communities for which we can reject the null in favor of the one sided alternative hypothesis $p^{alternative} > p^{null}$ together contain at least 200 eligible households.

Constructing the Lower Tail Sample To construct the *lower tail sample* we employed a similar strategy. First, we assume that in the absence of sex-selective abortions, 105 boys are born for every 100 girls. Consequently our null hypothesis is ($p^{null} = \frac{105}{105+100} \approx 0.512$). As an alternative hypothesis, we assume that sex-selective abortions are practiced in a given community, and we should therefore expect $x > 105$ boys to be born for every 100 girls ($p^{alternative} = \frac{x}{x+100}$). The lower tail sample consist of all communities for which we cannot reject the null hypothesis in favor of a given alternative hypothesis provided that we have adequate statistical power to actually reject the null hypothesis if the alternative hypothesis was true. The task was then to find the lowest alternative hypothesis $p^{alternative}$ such that the number of eligible households in the selected communities is at least 200.

Logistics of data collection To actually implement the implicit association test and the computerized questionnaire, we hired and trained approximately 20 facilitators and ran the software packages Inquisit (www.millisecond.com) and Limesurvey on laptops. The facilitators were recruited by the national statistical agency of Armenia, and most facilitators had prior experience with face-to-face data collection. We spent several days training these facilitators, and we practiced and refined our methods by running several pilot sessions with a sample of households that are not in our main study. After development and pre-testing, facilitators were divided into teams of three, and each team member was assigned a computer. When data was collected in one of the households in our study, the team would set up the three computers in different corners of the apartment and use folding dividers made out of cardboard to create a kind of isolation booth around each computer and participant.

To begin with a participant, a facilitator would sit down with the participant and explain the abbreviated implicit association test. The two of them would go through the test together, and the facilitator would evaluate the participant's understanding. They repeated this exercise as necessary. When the participant was prepared to proceed to the full implicit association test, the facilitator would help the participant put on headphones and then start the test. After verifying that the participant could hear the audio recordings, the facilitator would step away immediately, so that the participant could complete the test in privacy.

Most participants were assigned a facilitator of the same sex. While we would have liked to randomize participant-facilitator combinations within a given household, cultural norms around gender implied that it would have been difficult in some households for male facilitators to interview wives and for female facilitators to interview husbands. Faced with this trade-off, we allowed facilitators to exercise their own judgment regarding the best assignment of participants to facilitators within a given household, and we made sure that all data collection teams included both male and female facilitators.

Quality control. The data collection was monitored by independent employees of the *Women's Resource Center of Armenia*, who made both announced and unannounced visits to the teams of facilitators in the field. At least once a week the data was transferred to us so that we could check it for consistency. Thanks to log files stored on the computers, we were able to confirm, among other things, that i) the husband and the wife in a given household were interviewed at the same time by different facilitators ii) participants were allowed to proceed to the main implicit association test only once they had reached a threshold level of proficiency in the trial categorization task, or iii) enough time was spent on each questionnaire item for the participants to understand the question and think about how to answer it.

A.2 Conditional sex ratios

In figure A.2, we replicate figure 1 from the main text separately for the regions of Yerevan, Gegharkunik, and Syunik. We find extremely similar patterns in all three regions.

In figure A.3, we further dis-aggregate the sex ratio among third-born children. In particular, we compute the sex ratio at third parity separately for all possible gender compositions among the first two children. We find that in families with a mixed gender composition the point estimates for the sex ratio among third-born children lie above the natural rate of 105 boys for every 100 girls, suggesting that at least for some parents son bias goes beyond wanting to have at least one son. The degree of distortion in the sex ratio is much less severe, however, compared to families that were still lacking a son and not statistically significant at standard significance levels.

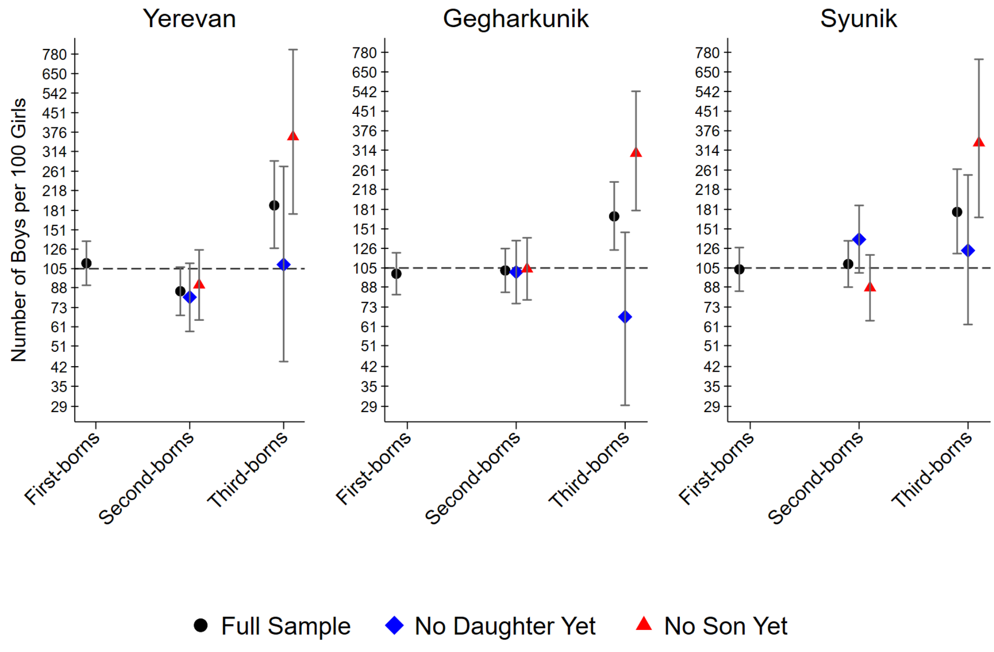

Figure A.2: Conditional sex ratios by region

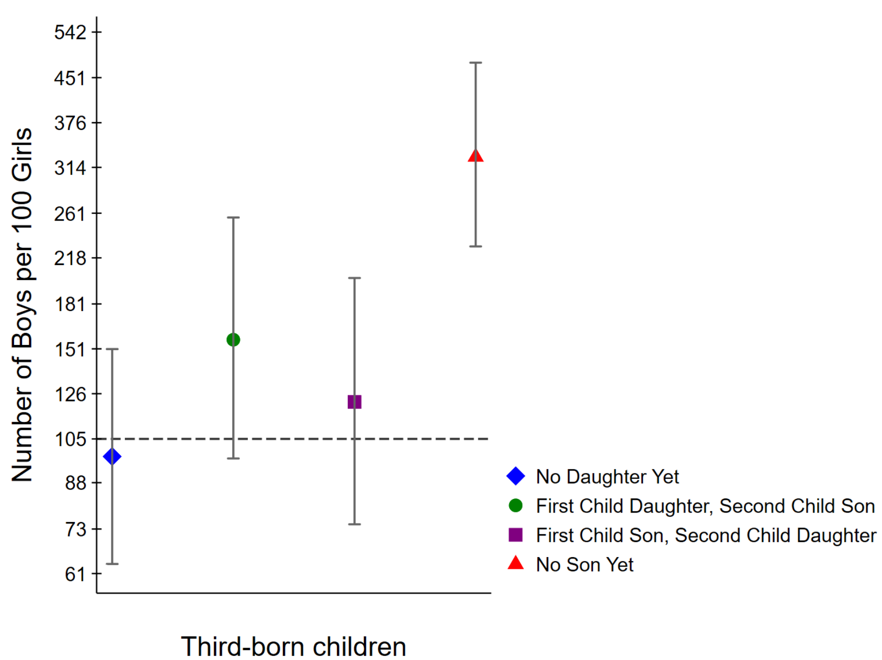

Figure A.3: The ratio of boys to girls among third born-children. 95% confidence intervals are computed using the Clopper-Pearson method for calculating binomial proportion confidence intervals.

Figure A.4: Sample drawings for family with two sons

A.3 Implicit Association Tests

Implicit association tests measure associations between target stimuli presented in neutral terms and valued stimuli. In our case, the target stimuli were drawings of families with either sons or daughters, while the valued stimuli were audio recordings of value-laden words.

Implicit association tests have been successfully used to study associations related to race, sexual orientation, religion, and other sensitive topics in contemporary societies [Nosek et al., 2007]. Implicit association tests are much less prone to producing socially desirable responses than traditional survey methods. We implemented our implicit association test with all of our sampled participants. Our implicit association test followed the structure presented in Nosek et al. [2007]. The test required participants to categorize all stimuli correctly. If a participant made a mistake, a large red “X” appeared on the screen, and the participant had to try again before continuing.

Stimuli. For our target stimuli, we presented drawings showing two different Armenian families engaged in daily activities like eating, watching television, or doing homework (see figures A.4 and A.5). In both families there are two children, but in one family the children are girls, and in the other family they are boys. A given drawing showed one of the two families engaged in one of the various activities.

The valued stimuli differed across the two versions of the implicit association test that we implemented. In the *Valence IAT*, participants were presented with positive and negative words. To positive category included the words “to cherish”, “spectacular”, “joyous”, “excellent”, and “glad”. The negative category included the words “yucky”, “horrible”, “to fail”, “harmful”, and “sad.” In the *Stereotype IAT*, participants were asked to sort words into two different explicitly labeled categories. The first category was labeled “to flourish” and included the words “offspring”, “descendant”, “to multiply”, “standing”, and “to immortalize”. The second category was labeled “to wither” and included the words “childless”, “infertility”, “to interrupt”, “to fade”, and “extinction.”

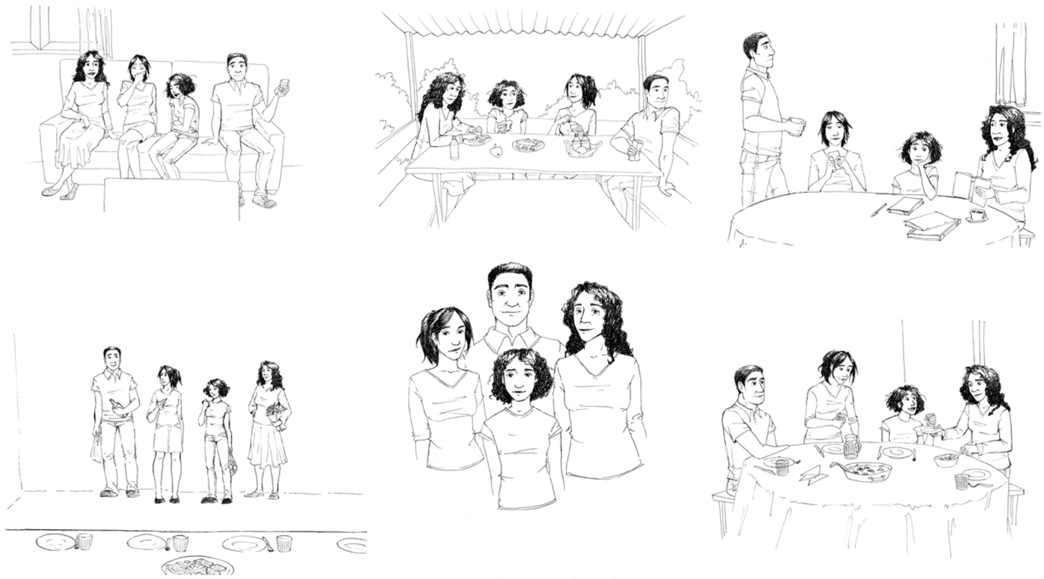

Figure A.5: Sample drawings for family with two daughters

Introduction of the families and counterbalancing scheme. At the beginning of the implicit association test, the two families were shown together on the screen, and participants listened to a short audio recording over headphones that introduced the families. All family members were introduced by their names, and the participants learned about their respective ages. We chose ages for the parents and the children that made it look very unlikely that another child would be born into this family in the future, and the participants were told explicitly that the parents would not have any more children.

To ensure that our stimuli are focused on the gender of the children rather than any attributes of the parents, we used a counterbalancing scheme in which half of our participants took a version of the implicit association test in which the family with the two daughters (sons) was headed by the first (second) set of parents, while the other half took a version in which the family with the two daughters (sons) was headed by the second (first) set of parents. Furthermore, the left and right positions of the two families on the introductory screen were also counterbalanced across participants. For the case in which the family with the two girls is on the left, the introductory recording translates as follows:

> Look at the two families on the screen.
>
> Look at the family on the left. This is the Hovhannisyan family. The husband, Tigran, is 45, and the wife, Hasmik, is 43. They planned to have two children. They have two daughters, and they will not have any more children. The oldest daughter, Nare, is 18 years old, and the youngest daughter, Marie, is 15 years old.
>
> Look at the family on the right. This is the Gasparyan family. The husband, Hayk, is 45, and the wife, Gayane, is 43. They planned to have two children. They have two sons, and they will not have any more children. The oldest son, Davit, is 18 years old, and the youngest son, Narek, is 15 years old.
>
> In the coming task you will see several images of the two families. Look at the two families closely and try to remember them. This exercise will take about 15 minutes. Press the long key with the white sticker at the bottom of the keyboard when you are ready to continue.

Following Greenwald et al. [2003], our implicit association test consisted of seven basic blocks of trials. We counterbalanced the order of categorization rules across participants. Specifically, some participants initially faced a categorization scheme that paired negative words with the family with daughters, and they later faced a scheme that paired positive words with this family. Counterbalanced participants went in the opposite order. Altogether, our complete counterbalancing scheme produced multiple versions of the test. Counterbalancing allowed us to distribute the versions evenly and unsystematically across participants, and this eliminated the potential for any artifacts associated with the parents, the spatial locations of stimuli, or the ordering of categorization rules.

Scoring algorithm. *D* scores quantify a participant's relative response times under the two categorization schemes in the implicit association test. To the extent that relative response times vary systematically with some underlying attributes of the participants that are of interest to the researcher – in our case the degree of son bias in our participants' fertility preferences – *D* scores can be interpreted as a measure of implicit associations and preferences. Several extraneous factors have been shown to contaminate the measurement of implicit associations by exerting an independent influence on *D* scores. These nuisance factors include the relative order in which the categorization tasks are presented and individual differences in average response latency [Greenwald and Nosek, 2001, Nosek et al., 2005, Greenwald et al., 2003, McFarland and Crouch, 2002].

We addressed these challenges in several ways. First, we randomized across participants the order in which the target stimuli were presented. Second, we familiarized all participants with the general protocol of the implicit association test by running a trial version that involved unrelated stimuli. Participants were allowed to proceed to the main test only once they had reached a threshold level of proficiency in the trial categorization task. Finally, we implemented the scoring algorithm developed in [Greenwald et al., 2003] that has been shown to minimize the influence of individual differences in average response latency.

This scoring algorithm produces *D* scores that are distributed on the interval [-2; 2] where a value of zero implies the absence of the implicit association in question. Positive scores in the Valence IAT imply an association of families with sons and words with positive valence. Similarly, positive scores in the Stereotype IAT imply an association of families with sons and words with connotations of flourishing.

Households were randomly assigned either the Valence IAT or the Stereotype IAT and all families members within a household completed the same version of the implicit association test. Because both versions produce similar results (cf. figure 3 in the main text), we combine the *D* scores from the Valence and Stereotype IAT for the main analyses in the paper. In all our analyses, we account for the versions of the IAT and the initial pairing of the target stimuli.

A.4 A Measure of Explicit Son-Biased Fertility Preferences

Husbands and wives in our sample are asked to envision the optimal future family of their youngest child, and we elicit preferences regarding the sex composition of this child's future children. Similarly, we elicit the preferences of the husband's mother regarding the same child's future children by adjusting the wording of the question to refer to this particular grandchild of hers. A first question asks for the optimal family size without giving the respondents the chance to specify the sex of the children. Then, at later points in the questionnaire, the respondents are asked to assume that the focal child would have exactly one, two, three, or four children respectively. In each of these four cases, the respondents are asked to specify the optimal number of boys and girls as well as the exact birth ordering while taking the overall number of children as given. To avoid anchoring effects, these four questions are each placed in different sections of the questionnaire, and the order in which the questions are presented is randomized across participants. Finally, in a last question, our facilitators present the respondents with four different scenarios regarding the future children of their youngest child and ask the respondents to pick their favorite scenario. These four scenarios describe families ranging in size from one to four children, and the questionnaire is programmed such that the gender distribution among the children corresponds exactly to the preferred gender composition of the participants as elicited in the earlier questions. Comparing the number of children in a participant's favorite scenario to the preferred number of children when the gender of the children could not be specified allows us to disentangle preferences over the number of children from preferences over the sex distribution.

Figure A.6 shows the relative frequency of reported optimal sex compositions for families of varying sizes. Two interesting features of the data stand out. First, there is a remarkable level of similarity in the relative frequency of preferred sex compositions across husbands, wives, and mothers-in-law. Second, the data suggest that there is not only a preference for having at least one son, but also a clear preference for a balanced sex composition of children. In fact, for the questions where the overall number of children is set to be two or four and the participants can therefore in principle achieve a balanced gender composition, more than 90% percent of all reported preferred sex compositions are completely balanced. However, despite the overall similarity in the distribution of preferred sex compositions across husbands, wives, and mothers-in-law, it is still possible to detect systematic gender differences. For example, the share of husbands that favor a family composition with more male than female children is higher than the share of women who favor such a family composition. Similarly, a higher share of husbands than of wives want the first child to be a son or want all children to be boys.

To summarize the degree of son-bias in the stated preferences in a single statistic, we collapse the information contained in figure A.6 into a single index of son bias. More specifically, we implement a scoring algorithm that assigns one point whenever the respondent (a) wants the first child to be a boy, (b) wants more boys than girls, or (c) wants all children to be boys. Importantly however, the index does not double count in cases where two scoring criteria are equivalent. In the case of one child, we have that (a), (b), and (c) are all equivalent, and hence participants can score at most one point, while in the case of two children we have (b) and (c) are equivalent and therefore participants can score at most two points. Summing over the four questions with varying fixed family sizes, the resulting index can take any integer between zero and nine[37]. Figure A.7 shows a histogram of our index of son bias by region and subject type.

A.5 Internal consistency of D scores and Index values

In this section, we show that the correlation between D scores and index values depicted in figure 4 remains positive and significant after accounting for potential confounding factors.

As is shown section IV.1 in the main text, the largest differences in D scores occur between male and female participants, and between participants living in different regions. A possible concern is therefore that i) our measures of son bias are directly affected by a participant's gender even if latent son bias is held constant, or that ii) there are unobserved regional variables that do not reflect son bias but affect our measures of son bias. To the extent that this is the case, the correlation documented in

[37]This index was preregistered by Efferson et al. [2018]

(a) Preferred sex of only child

(b) Preferred gender composition for family with two children

(c) Preferred gender composition for family with three children

(d) Preferred gender composition for family with four children

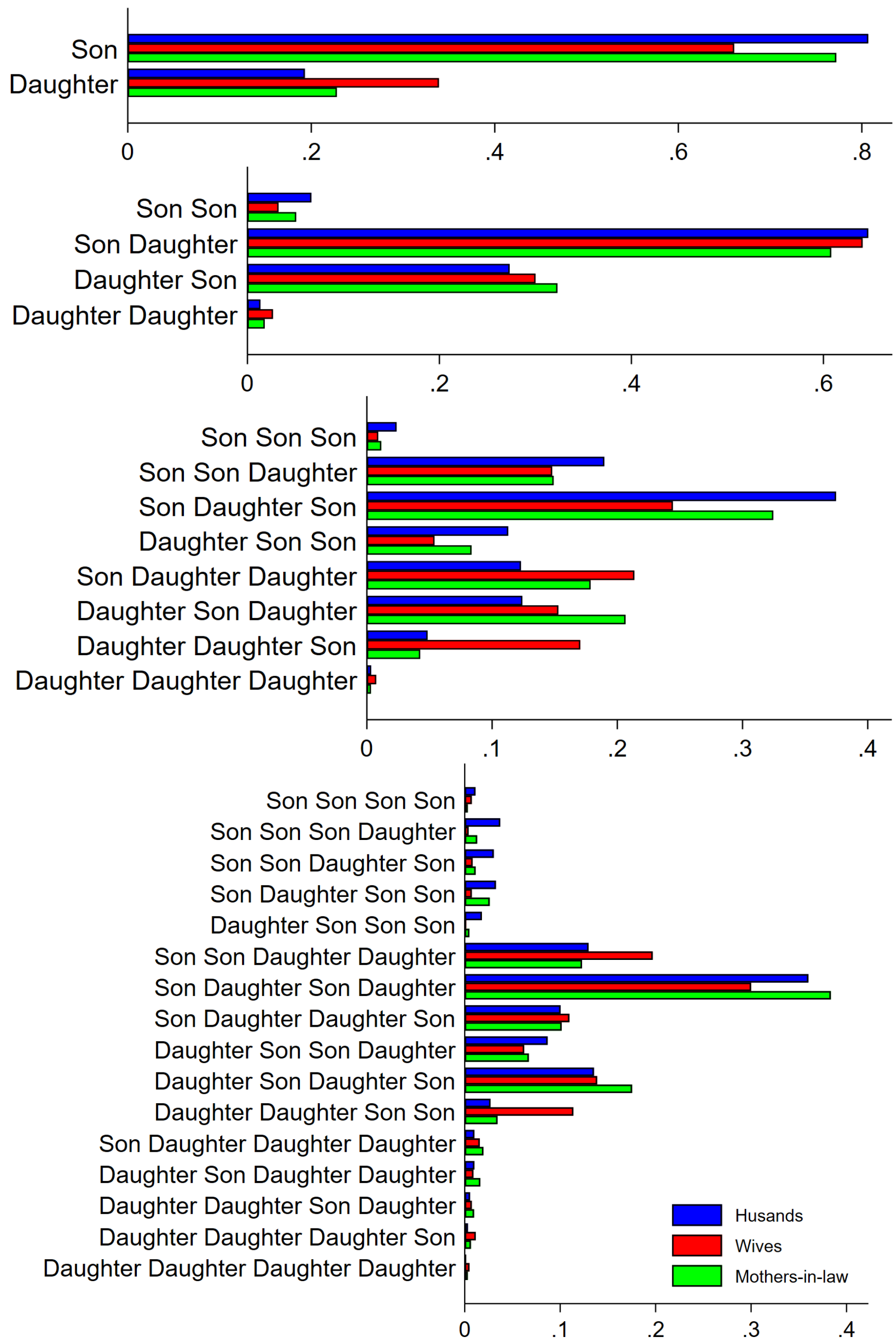

Figure A.6: Frequencies of reported optimal gender composition for the future families of own offspring.

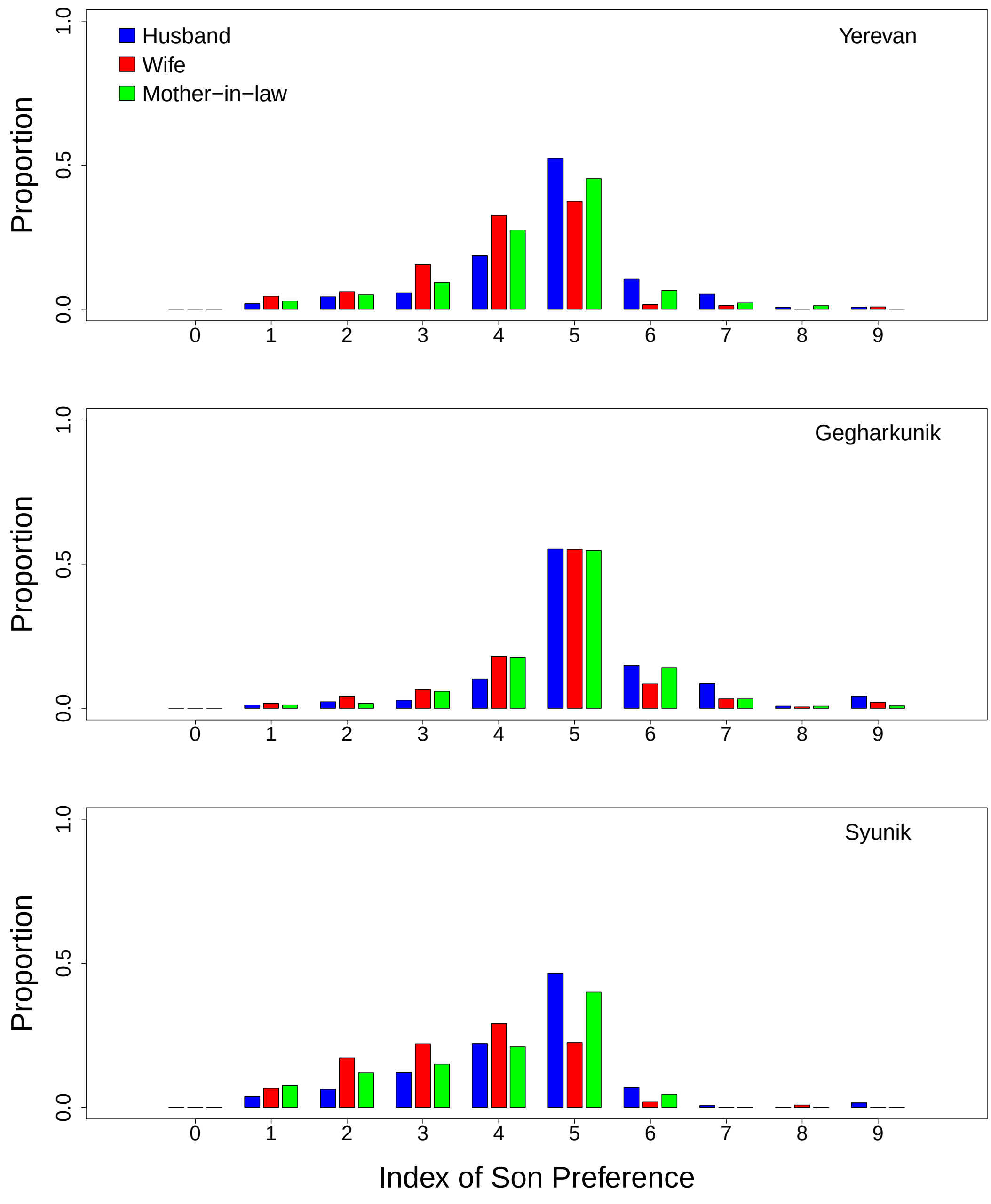

Figure A.7: Histogram of our index of son bias by region and subject type.

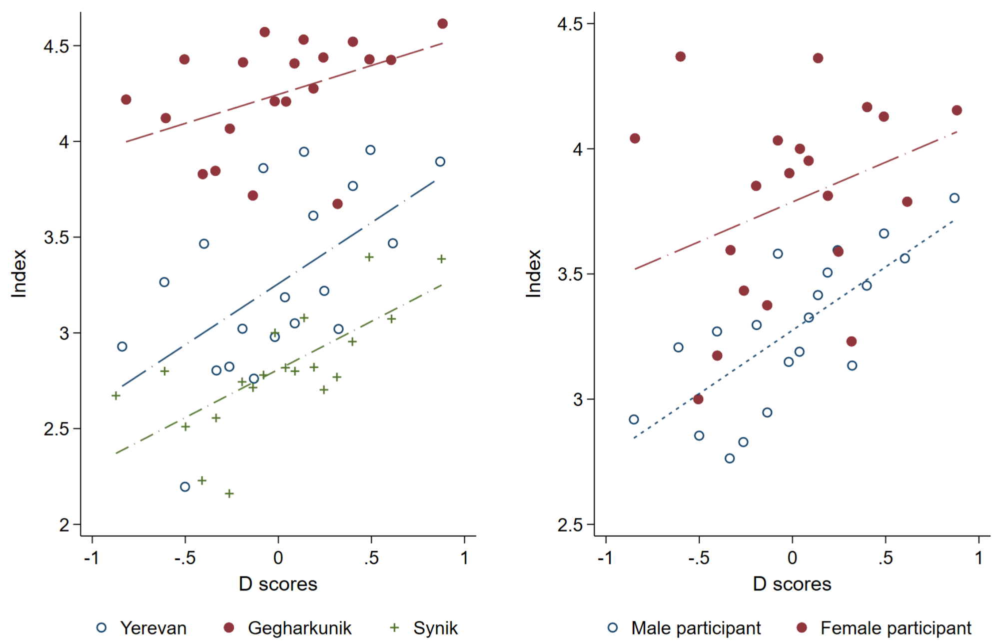

Figure A.8: Binned scatter plots

figure 4 would be spurious. Figure A.8 shows binned scatter plots of the associations between D scores and index values by gender of the respondent and by region. We find that the positive association between D scores and index values is also obtained *within* regions and *within* participants of the same gender.

This result can be put to a more rigorous test. Table 4 reports estimation results from a regression of index values on D scores. Column 1 documents the raw association between D scores and index values and establishes that higher D scores indeed predict higher values of the index. In column 2 we control for subject type and find that the association between explicit and implicit measures of son bias is robust to exploiting only variation in son bias within wives, husbands, and mothers-in-law. One may be concerned that both the D scores and the index are affected by the gender composition of the children in a participant's family. For example, similarity between one's own family and the families depicted in the implicit association test may affect response times. This will affect D scores. Similarly, respondents may take into account the gender of their youngest child when deciding on the ideal gender composition in this child's future family. This will in turn affect the index. This can lead to spurious correlations between the implicit and the explicit measures of son bias across participants with different family compositions. Column 3 therefore includes a full set of family composition fixed effects and shows that this potential confounding factor is not driving the results. In column 4, we additionally account for 45 community fixed effects. The coefficient retains its significance at the 5% level.

Given that D scores as well as index values are measuring true underlying son bias with some amount of noise, the estimated coefficient will suffer attenuation bias and we should expect the coefficient to become smaller as we move from column (1) to column (4). Controlling for region, subject type, or family composition absorbs some of the variation in son bias, as can be seen by the fact

	Son Pref. Index				
OLS estimates	(1)	(2)	(3)	(4)	(5)
D score	0.578***	0.426***	0.381***	0.205**	0.426***
	(0.0807)	(0.0845)	(0.0849)	(0.0795)	(0.129)
Subject type fixed effects	No	Yes	Yes	Yes	No
Family composition fixed effects	No	No	Yes	Yes	No
Community fixed effects	No	No	No	Yes	No
Household fixed effects	No	No	No	No	Yes
Observations	2695	2695	2695	2695	2695
Adjusted R^2	0.016	0.033	0.076	0.192	0.251

Standard errors are clustered at the household level.

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

Table 4: Internal consistency of our implicit and explicit measures of son bias.

that the adjusted R^2 increases from 0.016 to 0.192. The control variables cannot, however, absorb any measurement error. Hence, the signal-to-noise ratio worsens and attenuation bias becomes more severe.

Finally, in column 5 we include a full set of 1,212 household fixed effects. We find that even if we effectively control for any observed or unobserved variable that varies at the household level, the two measures of son bias are still strongly positively correlated with each other, and the association remains statistically significant at the 1% level. The coefficient in column (5) is identified from variation in D scores and index values across subject types and attests to the fact that measured son bias is systematically larger for men in our sample relative to women.

Comparison to standard measure of son bias A complementary exercise is to compare our measures of son bias to the more standard measure of son bias in the *Demographic and Health Surveys* (DHS). While we cannot replicate the DHS question using our questionnaire items, we can approximate it. For each participant, we can focus on the preferred scenario in terms of family size and compute the share of sons in this scenario. The resulting variable can take values between zero and one, with intermediate steps that depend on the family size in the preferred scenario.

The correlation between our index of son bias and this variable is 0.549. Because our index is based on the same underlying data, we should expect a positive correlation. Finding a correlation well below unity suggests that the index contains a considerable amount of information that is not contained in this simpler DHS-style variable. The raw correlation between our D scores and the DHS-style variable is 0.119. Interestingly, if we use the index and the DHS-style variable to predict D scores in a multivariate regression while accounting for subject-type fixed effects, region fixed effects, and own children fixed effects, only the index comes in significant. This suggests that our index contains more information or is less biased than the variable approximating the DHS-style question.

A.6 Construction of income, education, and occupation variables

The variables on income, education, and occupation used in table 2 in the main text are constructed as follows.

Income level Income levels are measured on an index scale from one to six, based on answers to the question (cf.[Khachatryan, 2019]):

> "Which of the following statements best describes the financial situation of your household?"

Answer categories are as follows (percentages of answer in parentheses).

1. The income is hardly enough to buy food (35.8%)
2. The income is enough to buy food, but we can't afford new clothes (23.7%)
3. The income is enough to buy new clothes, but we can't buy technical equipment (27.4%)
4. The income is enough to buy technical equipment, but we can't buy a new car (8.8%)
5. The income is enough to buy anything but a new apartment (3.7%)
6. There are no financial difficulties, and we could buy an apartment if needed (0.6%)

We also asked participants about the monthly household income in Armenian Dram. The index defined above is strongly correlated with self-reported household income ($\rho = 0.44$). However, the correlation is stronger for men ($\rho = 0.48$) than for women ($\rho = 0.39$), which is expected if men in our sample are more likely than women to be the primary breadwinners and to have more accurate information about the exact level of household income. We prefer the index as our main measure of household income because it is likely to be relatively comparable across subject type.

Education level Education levels are measured on an index scale from one to six, based on the question:

> "What is the highest level of education that you have completed?"

Answer categories are (percentages of answer in parentheses):

1. Primary school or no schooling completed (0.6%)
2. Middle school (3.6%)
3. High school (42.2%)
4. Vocational education (27.2%)
5. Higher education (25.3%)
6. Post graduate education (1%)

Occupation We asked husbands and wives about their current occupations. Since we expected a large fraction of mothers-in-law to have reached retirement age, we instead asked them about their main occupation in their thirties and forties.

	Share of sons			Last child son		
Mother's degree of son bias	0.096***		0.078***	0.128***		0.110***
	(0.016)		(0.016)	(0.025)		(0.025)
Father's degree of son bias		0.100***	0.084***		0.105***	0.083***
		(0.015)	(0.015)		(0.024)	(0.025)
R-squared	0.046	0.054	0.082	0.032	0.022	0.045
N	766	766	766	766	766	766

Table 5: OLS regresssions of son-biased fertility outcomes on a measure of parental son bias. The measure of son bias used in these regression is the composite measure described in section IV.2. We include in these regressions all households with more than one child in which we were able to collect data from both the husband and the wife.

A.7 Fertility outcomes and measured son bias of husbands and wives

Whether the fertility preferences of father or mothers bear more responsibility for the skewed sex ratio in Armenia is an important question. Table5 shows that fertility outcomes at the household level are correlated with measured son bias of both mothers and fathers, suggesting that the fertility preferences of both parents matter. Note, however, that these individual-level correlations reflect not only the effect of parental son bias on fertility outcomes, but are likely also partly driven by the causal effect of realized fertility outcomes on measured son bias (cf. III.3 in the main text).