# Money, Time, and the Demand for Children

Evie Thompson (LSE) Guy Michaels (LSE)*

November 2, 2026

Click here for the latest version

US fertility has fallen well below replacement, even as stated fertility ideals have not. We ask how much money and time women are willing to trade off for additional children, a question that is difficult to answer using observational data or even natural experiments. We field a discrete choice experiment in which 1,995 women choose between hypothetical futures varying the number of additional children, household income, and time use. Demand falls with both income and time prices, and rises with income faster than proportionally, consistent with larger families being a luxury good. Exploratory results show that full-time work, but not part-time work, is perceived as costly, and that baseline demand varies by age, religiosity, and desired fertility even as price responses vary less. Notably, women on average require compensation, often significant, to accept a second additional child, including many who say they want two more. This finding complicates the standard reading of a stated ideal of two children as evidence of unmet demand: once its price is held fixed, we find little demand to reach it.

Keywords: Fertility, Gender, Discrete Choice Experiment, Time Use, Willingness to Pay
JEL codes: J13, J16, J22

*For helpful comments, we thank many colleagues, especially Tim Besley, Matthias Doepke, Steve Pischke, and Carmit Segal, and workshop participants at LSE. We also thank the team at Bilendi for their help with the experiment. This research is funded by the research allocation of Guy Michaels at LSE. The study has LSE research ethics approval (reference 655943). This study was pre-registered with the AEA RCT Registry (AEARCTR-0019209) on July 20, 2026, prior to the collection of confirmatory data starting July 22, 2026; see `https://doi.org/10.1257/rct.19209-1.1`. Evie Thompson's surname was Jamieson when the study was registered and run, so her former name appears in the registry and on the consent form. Permission to deposit anonymized responses in a data archive was part of the consent required to participate; anonymized data and code will be made available upon publication.

## 1 Introduction

The US total fertility rate (TFR) fell in recent decades to about 1.6 births per woman in 2024, well below its replacement level. Meanwhile, global TFR has fallen from over 5 in the 1960s to around replacement level or possibly even below it (Hamilton et al., 2025; United Nations, Department of Economic and Social Affairs, Population Division, 2025; Fernández-Villaverde and Norrick, 2026). These fertility trends have economic and fiscal consequences, through their impact on population aging, labor supply, healthcare costs, and the financing of social-insurance systems. Recent work studies the drivers of low fertility in high-income countries, emphasizing factors such as the compatibility of work and family life, intensive parenting expectations, and the opportunity cost of parents' time and resources (Doepke et al., 2023; Kearney and Levine, 2025). These costs fall disproportionately on women, with much of the earnings gap between men and women opening up after a first birth (Cortés and Pan, 2023). In this paper, we study women's willingness to trade off money and time for additional children, using an experiment that reveals these tradeoffs.

Classical and modern fertility theory alike treats children as chosen against money and time shadow prices (Becker, 1960; Mincer, 1963; Willis, 1973; Galor and Weil, 1996; Doepke et al., 2023; Delventhal et al., 2021). But those prices have never been recorded directly: existing estimates derive them theoretically or infer them indirectly from realized fertility using structural methods (Hotz and Miller, 1988; Adda et al., 2017). Even natural experiments do not suffice, since realized fertility may reflect preferences of partners (Doepke and Kindermann, 2019) and family, unplanned pregnancies, and fertility problems. Moreover, women systematically fail to anticipate the employment consequences of motherhood when making earlier life decisions (Kuziemko et al., 2018). Closing these gaps requires observing choices made under explicit money and time prices, which is what our experiment does.

Demographic surveys measure fertility preferences by asking about ideal or intended family size. They have long documented an ideal of around two children, even as realized fertility has fallen below it (Hagewen and Morgan, 2005; Sobotka and Beaujouan, 2014).[1] This gap is commonly interpreted as evidence of unmet demand, whereby constraints prevent women from realizing their fertility intentions. But these surveys do not ask what a woman would give up, in money or time, to realize those intentions. Existing fertility experiments either weigh money and time against family size qualitatively (e.g., Sheppard, 2025; Behrman et al., 2025) or elicit a desired number of children under randomized allowances, childcare costs, and wages (Goldstein et al., 2026); neither prices an additional child in hours and dollars. In economics, stated-preference choice experiments have been used to recover money-metric valuations of other major life and career decisions, including job attributes and the career-family trade-off itself (Wiswall and Zafar, 2018, 2021); we apply this approach to the demand for children.

We develop and implement a stated-preference experiment to study demand for children, asking

[1] As we show in the appendix, fertility intentions reveal a similar gap. Using the National Survey of Family Growth, we compare intentions at ages 20 to 24 with completed fertility at ages 40 to 44 for four cohorts born between 1958 and 1982. Intended family size exceeds completed fertility in every cohort; across the four, the gap averages 0.34 children.

how women trade additional children against household income and their own time. We characterize the shape and heterogeneity of this demand: whether the time cost rises smoothly with hours or is concentrated in particular uses of time, whether a second additional child is valued differently from a first, and how valuations vary with household income and across women. We also compare these contextualized choices with women's stated fertility ideals.

The experiment presents each of 1,995 women aged 23–35 with ten hypothetical choices between two descriptions of their life ten years hence. The alternatives independently vary the number of additional children, combined pre-tax annual household income, and the allocation of a typical weekday across paid work, care and household tasks, and leisure and sleep. Because typical household incomes differ substantially by education, the income levels shown are stratified by whether the respondent has a college degree. We focus on income and time for three reasons: they are the central resource-allocation margins in classical models of fertility (Becker, 1960; Becker and Lewis, 1973; Willis, 1973); the child-penalty literature documents substantial costs to women through both channels (Kleven et al., 2019a, 2025; Cortés and Pan, 2023); and family policies directly affect households' financial resources and the time costs of raising children, a cost the literature has so far inferred only indirectly, through variation in childcare prices, availability, or leave policy, rather than elicited directly (Doepke et al., 2023; Bergsvik et al., 2021). Our design allows us to estimate separately how women trade an additional child against the different attributes. Because each respondent answers ten such choices, we estimate these trade-offs within person, and separately for women who are currently partnered and those who are not. These estimates isolate a woman's own demand for children in a future with a partner from her partner's preferences and the other factors that shape realized fertility.

Confirmatory analysis shows that demand for additional children declines as either its money or time price rises: a 10 percent reduction in the household income attached to the larger-family scenario lowers the probability of choosing it by 2.8 percentage points, and a reduction of one hour of leisure per weekday lowers it by 0.6 percentage points. Higher-income scenarios result in a greater share of income foregone for children, consistent with larger families being a luxury good: the value of an additional child rises with scenario income faster than in proportion ($\hat{\mu} = 0.028$, 95% CI $[0.012, 0.045]$).

Exploratory analyses that extend beyond our pre-registration reveal further detail. Care hours reduce demand linearly, but paid work does not: part-time work does not reduce demand relative to no work, while moving from part-time to full-time work reduces it sharply, requiring compensation of about 20 percent of household income. Baseline demand varies substantially by religiosity, desired fertility, and age. For example, holding desired fertility and age fixed, moving from the least to the most religious respondents is associated with a 19.0 percentage-point increase in the probability of choosing the larger family, and each additional desired child with a 7.9 percentage-point increase. In contrast, price responses vary less, although women who want more children respond more to the money price and more religious women respond less to the time price.

Our findings are not artifacts of task order, of respondents who simply default to the higher-

income option regardless of price, or of our functional-form and model choices, although the time-price response strengthens in later tasks. We also address potential concerns that shown attributes can shift beliefs about unshown ones, such as a partner or a job, akin to omitted-variable bias (Adrjan et al., 2026). Few open-text comments show such inferences, and partnered and unpartnered respondents, who differ in how much they must infer a partner, show little difference in the money-price estimate. This reflects our decision not to assume a unitary household: we leave how income and care are split between partners unobserved, rather than imposed.

Although the experimental choices are hypothetical, our estimates are broadly consistent in direction with independent, real-world evidence from other countries and other sources of variation, and in magnitude where the external estimate is built on a comparable horizon. Converting each external shock into our units assumes, among other things, that a change in the probability of choosing a scenario maps one-for-one, in percentage points, onto a change in the probability of a birth, so we treat these as order-of-magnitude checks. Our predicted response to Spain's one-off transfer is of the same order of magnitude as the published estimates (González, 2013; González and Trommlerová, 2023), as is the estimate built on a full year of Poland's Family 500+ (Bokun, 2024); a shorter, announcement-only window recovers less (Gromadzki, 2024). A randomized Swedish lottery win provides a less-powered check on the luxury-good result: our predicted effect is consistent with the study's estimated range (Cesarini et al., 2023). Our exploratory finding that religiosity predicts demand is also consistent with the completed-fertility gap by religiosity in the National Survey of Family Growth (NSFG) (Hayford and Morgan, 2008).

Strikingly, our exploratory analysis shows that the value of an additional child falls sharply from the first to the second: at the median income of each education group, non-college women would give up about $4,600 for a first additional child (not statistically distinguishable from zero) and college women about $20,200, but they would require about $26,300 and $29,300, respectively, to accept a second. The willingness to pay for a second additional child is negative even among women who say they want two more.

A stated ideal may be an aspiration rather than a priced choice: it need not describe what a woman would choose once that choice carries explicit consequences for income and time. Our experiment tests this directly, comparing respondents' ideal family size with their choices between such concrete futures. Ideals correlate strongly with these choices but do not translate into them one-for-one. In our registered exploratory test, childless women who want none adhere to that ideal far more often than those who want one; the latter often choose no children instead, showing that ideals overstate demand for this group.

Our results on how full-time work lowers demand for children also have implications for the standard economic model of the time cost of children. Becker's price-of-time framework values each hour taken from paid work at the wage, implying a cost that rises smoothly with hours (Becker, 1965); our results depart from this, with the cost of paid work concentrated almost entirely in the transition to full-time employment rather than rising with hours worked. This pattern is more consistent with the inflexibility of full-time jobs themselves (Goldin, 2021; Mas and Pallais, 2017)

than with a continuously wage-priced value of time, adding to the time-cost explanation emphasized by Doepke et al. (2023). The child-penalty literature documents the same conflict between full-time work and motherhood after a first birth, when many mothers reduce their hours or stop working (Kleven et al., 2019a; Cortés and Pan, 2023). Our results show that full-time work is costly in women's choices even before any birth, consistent with a convexity Goldin (2014) documents in wages.

The remainder of the paper proceeds as follows. Section 2 describes the experimental design and data. Section 3 sets out the estimation approach. Section 4 reports the results. Section 5 compares our estimates to external evidence.

## 2 Experimental design and data

### 2.1 Experimental design

#### *2.1.1 Overview of the choice experiment*

We use a survey experiment to elicit women's demand for additional children at different levels of household income and allocations of their own time. Each respondent answers ten questions, a number chosen to balance within-person statistical power against the risk of fatigue-driven inattention over a longer sequence, and in line with the number of tasks used in comparable discrete-choice studies of job and family trade-offs (Wiswall and Zafar, 2018; Mas and Pallais, 2017). Each question asks her to choose between two possible lives ten years in the future. These scenarios differ in three attributes: the number of additional children she has, her combined household income, and how she divides a typical weekday between paid work, care and household tasks, and leisure and sleep. Varying all three independently lets us separate the effects of household income and time use on demand for additional children.

Fertility decisions depend on many things, such as a partner's preferences, housing, childcare arrangements, and employment conditions. We limited the experiment to three attributes to reduce cognitive load in an already complex decision. We selected income and time because the child-penalty literature documents substantial costs to women on both margins, through lower earnings and increased care work (Kleven et al., 2019a; Cortés and Pan, 2023). They are also key dimensions of the household resource-allocation problem in classic economic models of fertility, which treat family size as jointly determined with the allocation of income and time (Becker, 1960; Becker and Lewis, 1973; Willis, 1973). Finally, they correspond closely to the main policy levers: cash benefits and tax credits operate through household resources, while childcare and leave policies alter the time costs of raising children. We do not vary expenditure: respondents choose their own spending within these constraints, so we price the opportunity cost of an additional child, not a spending bundle. Goods costs are therefore paid out of the stated household income and are not subtracted from our estimates.

*2.1.2 Attributes, levels and calibration*

**Additional children.** The first attribute takes values of zero, one, or two additional children over the next ten years. As with our other design choices, this range and horizon trade off realism against budget, survey length, and respondent cognitive burden — a shorter horizon, or room for larger families, are equally reasonable choices we did not make. We chose ten years so that partnering, and potentially more than one birth, would not feel unrealistically rushed, and capped additional children at two because, given prevailing fertility rates, this is where most women's actual decisions lie; three or more matters for only a small share. The instructions define these as children the respondent would have during that period, "not counting any children you already have today." The distinction matters because some respondents are already mothers: for a childless woman, moving from zero to one additional child is a first-birth decision, whereas for a mother it represents a higher-order birth. The definition includes biological children, including through surrogacy, as well as stepchildren and adopted children.

**Household income.** The second attribute is the combined annual income of the respondent and her partner before tax, ten years in the future. Respondents are told that this level would be reached gradually over the ten-year period and would then remain stable, and are asked to ignore inflation. To give combined income a common meaning, we instruct every respondent to "assume that you will have a partner or spouse 10 years from now, regardless of whether you currently have one."

The partnership assumption standardizes the hypothetical household across respondents and across scenarios. Without it, a difference in combined income could also reflect a difference in whether the respondent has a partner, rather than a change in the resources available to a household with the same structure, and it lets combined income vary independently of the respondent's own paid hours, since a partner's earnings mean the two need not move together. We use combined income rather than the respondent's own earnings because household resources, however they are allocated between partners, are what a fertility decision in practice draws on (Doepke and Kindermann, 2019) — not because we assume a single joint household decision; Section 4.7 sets out what this assumption does and does not imply, and reports estimates separately for currently partnered and unpartnered respondents. The assumption is a feature of the hypothetical choice, not a prediction about any respondent's future relationship status; the experiment therefore elicits demand conditional on having a partner ten years ahead.

Before the experimental block, respondents answer a short set of demographic questions used to confirm eligibility and ascertain their education level. We condition the income levels shown in the experiment on education because it was the strongest predictor of household income ten years later among the baseline characteristics we examined in the Panel Study of Income Dynamics (PSID). We also considered age, metro residence, and parental education. Although metro residence predicted future income, the predominantly metropolitan sample left too few non-metro respondents to support separate income grids reliably. We ruled out parental education because many women in the PSID did not report their father's education, making it an unreliable basis for assigning survey

respondents to income grids. We therefore construct separate grids for college graduates and women without a college degree. This keeps the displayed incomes plausible for each group while preserving random assignment of income within each education group.

We derive these grids from the PSID, using women who were aged 23–35 at baseline and observed approximately ten years later. Because the experiment asks respondents to assume that they will have a partner at that point, the calibration sample is restricted to women who are partnered at the later observation. Within each education group, the four displayed levels (Table 1) are the weighted 25th, 50th, 75th, and 90th percentiles of combined income. The three quartile points describe the center and spread of the distribution, while the 90th percentile extends the upper end of the experimental support, giving us greater power to detect modest income responses.

**Time use.** The final attribute describes the respondent's own time use on a typical weekday ten years in the future, split into three categories: paid work; care and household tasks; and leisure and sleep. The baseline static labor-supply model treats time as paid work versus leisure alone, but home production has been central to the economics of the family at least since Becker (1965), so we add it as a second category rather than folding it into leisure. Keeping to three categories, this is the most natural split available; commuting is pooled with paid work, and sleep with leisure. We calibrate its levels using time diaries from the American Time Use Survey (Flood et al., 2025) for partnered women aged 33–45 observed on a non-holiday weekday. This is a contemporaneous cross-section of women at the ages our respondents will reach in ten years, rather than a panel following the same women over time.[2]

Paid work is shown at zero, four, or nine hours, the rounded median weekday hours, including commuting, among women who are not in the labor force and those working part-time and full-time, respectively. Care and household tasks are shown at two, three, five, or eight hours, the rounded 20th, 40th, 60th, and 80th percentiles of the care and household time distribution. Because care hours are varied independently of additional children, we use four evenly spaced percentiles to ensure that the experimental support covers plausible levels of care in futures both with and without additional children.

Leisure and sleep form the residual category, so each allocation sums to 24 hours. The three paid-work levels and four care and household levels initially define twelve weekday allocations, of which three leave fewer than 10 or more than 20 hours in this residual category. We exclude these to improve realism, leaving nine allocations with 10 to 19 hours assigned to leisure and sleep.

To assess whether respondents interpreted the scenarios as intended, we conducted twelve cognitive interviews in which participants described their thoughts aloud as they completed the survey. We used these interviews to clarify the attribute definitions and instructions. We then tested the revised instrument in a pilot survey of 153 respondents and used the pilot evidence to finalize the income grids and comprehension checks.

[2]The calibration pools survey years 2003–2019 and 2021–2024; the ATUS diary-day weight is unavailable in 2020. Restricting the sample to 2015–2019 and 2021–2024 produces the same rounded levels.

### *2.1.3 Experimental variation and task construction*

Within each scenario, the number of additional children, combined household income, and weekday allocation are varied independently, subject only to the restriction to the nine feasible weekday allocations described above. In particular, household income is assigned independently of the respondent's paid-work hours. This breaks the mechanical link between her earnings and hours that would arise if income were constructed from her own labor supply, allowing changes in household resources to be distinguished from changes in her time use.

We implement this variation using complete randomization. The resulting identification is design-based: over the support of the experiment, the average effects of each attribute are identified by random assignment and do not depend on a particular parametric model of utility or distributional assumption about unobserved utility. This leaves us free to estimate the experimental relationships flexibly.[3]

Within each education group, the attribute levels define 108 possible profiles: three child levels, four income levels, and nine weekday allocations. For each question, the two profiles are drawn independently and uniformly from this set. We reject and redraw only fully identical pairs, which offer no meaningful choice, and otherwise leave the draws unconstrained. This includes pairs that differ only in income. These are the only pairs containing a dominated profile under our maintained assumption that preferences are monotonic in income, and we retain them as a descriptive check on whether respondents behaved rationally when the choice was trivial.

### *2.1.4 Instrument and respondent understanding*

To ensure that respondents understood how to read the scenarios, the instructions were followed by a worked example and three comprehension questions, one concerning each attribute. The correct answer was shown after each response. On the first attempt, 83.8 percent answered the children question correctly, 75.5 percent answered the income question correctly and 39.9 percent answered the time-use question correctly. We use these responses to examine whether the results differ among respondents who initially understood all three attributes. The full questionnaire is reproduced in Appendix K.

The survey also included an attention check. After making their ten experimental choices, respondents reported how realistic they found the scenarios and described in an open-ended question what they had considered when making their choices. Together, these measures provide evidence on respondents' engagement with the experiment and their interpretation of the scenarios.

The next section describes the sample and survey fielding, the construction of the analysis sample, and the weights used in estimation.

[3] A Bayesian $D$-efficient design could concentrate choices where a specified model predicts the most information, but only if that model and its priors are correctly specified; our pilot-based power calculations showed complete randomization gave adequate precision for the pre-specified effects, so we used it instead.

## 2.2 Data

### *2.2.1 Sample and fieldwork*

We fielded the survey between 22 July and 17 August 2026 through Bilendi, an online panel whose members volunteer for research in return for points redeemable for cash or vouchers. The survey was open to women living in the United States who were born between 1990 and 2003. We recorded birth year rather than exact date of birth, which allowed us to determine eligibility while limiting the personal information collected. These cohorts span the target age range of 23–35 in 2026, although respondents in the two boundary cohorts may have been 22 or 36 when they completed the survey.

The primary sample comprises respondents who answered all ten choice questions and spent more than four minutes on the survey, a threshold fixed before main fieldwork. A further 299 respondents answered all ten questions within four minutes and were excluded under this rule. The resulting sample contains 1,995 respondents and 19,950 choices across 39,900 scenario profiles. These respondents received points worth approximately \$1.90–\$2.15. At a median completion time of 446 seconds, or 7.4 minutes, that is equivalent to approximately \$15.30–\$17.40 per hour.

The reduced-form estimates use the 13,442 tasks in which the two scenarios differ in the number of children; every respondent answered at least one. The structural estimates use all 19,950 tasks, because a choice between scenarios with the same number of children still reveals how respondents value income and time. Desired-fertility analyses are restricted to the 1,976 respondents who reported both their current and ideal number of children.

### *2.2.2 Sample composition and weighting*

Because Bilendi's panel is opt-in rather than a probability sample, the realized sample need not be representative of the target population. We therefore gave Bilendi target recruitment shares for five respondent characteristics that change relatively slowly over adult life (education, age, urban-rural status, race and ethnicity, and parenthood) before fieldwork began, and Table 2 compares the realized sample with the corresponding population figures. Education, age, race and ethnicity, and metropolitan status are benchmarked to the 2024 one-year American Community Survey, and parenthood to the June 2024 Current Population Survey Fertility Supplement, which records whether a woman has ever given birth rather than whether children currently live in her household. Four of the five margins fall within 1.7 percentage points of their benchmarks. The urban share is 5.5 percentage points lower, although this comparison is approximate because the ACS metropolitan measure is broader than the city-or-suburb classification used in the survey.

The primary estimates are unweighted and therefore describe average responses in the realized sample. As a robustness check, we reweight the sample using four cells formed by crossing college status with age at 30. The resulting weights range from 0.948 to 1.102. Their design effect is 1.003, giving an effective sample size of 1,988 respondents. The variance cost of weighting is therefore small.

#### *2.2.3 Response diagnostics*

The median respondent spent 446 seconds on the full survey and 109 seconds on the ten choice tasks. The attention check was answered correctly by 93.4 percent of respondents, and 24 respondents, or 1.2 percent, chose the scenario on the same side of the screen in all ten tasks. Randomization also produced tasks in which the two scenarios were identical in children and time use but differed in income. Under our maintained assumption that preferences are monotonic in income, the higher-income scenario in such a pair dominates, and choosing the lower-income scenario is an error. Among the 484 respondents who faced at least one such task, 44, or 9.1 percent, made this error at least once. We use these diagnostics to define the robustness samples in Section 4.6, re-estimating the confirmatory tests on respondents who pass each check.

Perceived realism addresses a separate question: how respondents interpreted the scenarios. Of the 1,994 respondents who answered it, 1,336 reported that at least a majority of the scenarios felt realistic.

## 3 Methodology

### 3.1 Models

We analyze the choice data with a sequence of models that impose progressively more structure. We first report the observed choice shares, imposing no functional form. We then use the randomized variation in income and time to estimate reduced-form price responses; because both prices are randomized within respondent, these responses are identified by the design rather than by a model of how choices are made. The reduced form imposes a parsimonious functional form, but not the random-utility structure needed to recover valuations. We add that structure to express the value of an additional child as a compensating variation, the income or weekday time that holds utility fixed across the change in family size. We compare the structural valuations with the reduced-form price responses and the observed choice shares, and find them consistent throughout.

#### *3.1.1 Money and time prices*

For every task in which the scenarios differ in additional children, let $m$ denote the scenario with more children and $f$ the scenario with fewer. We summarize the income and time tradeoffs between them using two relative prices.

The money price is the difference in log household income,

$$P_{it} = \ln Y_{f,it} - \ln Y_{m,it}. \tag{1}$$

It is positive when the larger-family scenario has lower household income and negative when it has higher household income. The log specification is consistent with the structural model and expresses income differences proportionally across the two income grids. Although the models are estimated using $P_{it}$, we present magnitudes as the corresponding share of income in $f$ forgone in $m$,

$1 - \exp(-P_{it})$.

The time price is

$$T_{it} = (W_{m,it} + H_{m,it}) - (W_{f,it} + H_{f,it}), \tag{2}$$

where $W$ and $H$ denote the respondent's paid-work and care and household-work hours on a typical weekday. Since leisure and sleep make up the remainder of the day, a positive value is the number of hours of leisure and sleep forgone in the larger-family scenario.

### *3.1.2 Choice shares*

Throughout the paper, we present observed choice shares before introducing a model. In the main demand analysis, we report the share choosing the larger-family scenario at each fielded money and time price, both overall and, for the money price, separately by family-size comparison. We also report the choice shares for each family-size comparison when the two scenarios have identical displayed resources.

### *3.1.3 Reduced-form demand*

We use a reduced-form model to estimate jointly how the probability of choosing the larger-family scenario responds to the money and time prices. On tasks in which the two scenarios differ in additional children, let $y_{it} = 1$ if respondent $i$ chooses the larger-family scenario in task $t$. We estimate the linear probability model

$$y_{it} = \delta_i + \theta P_{it} + \phi T_{it} + u_{it}, \tag{3}$$

where $\delta_i$ is a respondent fixed effect. The coefficient $\theta$ gives the change in the probability of choosing the larger-family scenario associated with an increase in its money price, while $\phi$ gives the change associated with an additional weekday hour taken from leisure and sleep.

The price coefficients are identified from changes in a respondent's choices as the randomized prices vary across her ten tasks. Respondent fixed effects absorb differences across women in their average propensity to choose the larger-family scenario. All specifications control for whether the larger-family scenario appears on the left. We report unweighted estimates and cluster standard errors by respondent.

The money price captures the proportional income difference between the two scenarios. To distinguish this price response from an income-level response, we ask whether the larger-family scenario becomes more attractive when both scenarios offer proportionally higher household income. We therefore allow the response to each additional child to vary linearly with log income. This requires a main effect for the difference in additional children,

$$\Delta k_{it} = k_{m,it} - k_{f,it}, \tag{4}$$

and a child-by-income interaction,

$$C_{it} = k_{m,it}\left(\ln Y_{m,it} - \overline{\ln Y}\right) - k_{f,it}\left(\ln Y_{f,it} - \overline{\ln Y}\right). \tag{5}$$

We estimate

$$y_{it} = \delta_i + \theta P_{it} + \phi T_{it} + \eta \Delta k_{it} + \mu C_{it} + u_{it}. \tag{6}$$

Here, $\overline{\ln Y}$ is mean log income across primary-sample scenarios. The coefficient $\eta$ measures the response to an additional child at mean income, while $\mu$ measures how that response varies with household income. Because increasing both scenario incomes by the same proportion leaves the money price unchanged, $\mu$ separates the income-level response from the response to the price of a larger family.

To distinguish between different demands on respondents' time, we also estimate a specification that replaces $T_{it}$ with separate differences in paid-work and care and household-work hours. A further specification allows these effects to vary across the fielded hour levels.

### *3.1.4 Structural model*

The reduced form shows how choices respond to the money and time prices. To estimate the income or time equivalent of an increase in family size, we need the relative utility weights respondents place on additional children, household income and their own time. We recover these weights using a random-utility model.

Let $U_{ijt}$ denote the utility respondent $i$ receives from scenario $j$ in task $t$. The primary specification is

$$\begin{aligned} U_{ijt} = {} & \alpha_1 \mathbf{1}\{k_{ijt} = 1\} + \alpha_2 \mathbf{1}\{k_{ijt} = 2\} + \beta_Y \ln Y_{ijt} + \beta_W W_{ijt} + \beta_H H_{ijt} \\ & + \mu k_{ijt}\left(\ln Y_{ijt} - \overline{\ln Y}\right) + \varepsilon_{ijt}, \end{aligned} \tag{7}$$

where $k_{ijt}$ is the number of additional children, $Y_{ijt}$ is household income, and $W_{ijt}$ and $H_{ijt}$ are paid-work and care and household-work hours. Zero additional children is the omitted category.

The specification carries forward the proportional money scale of the reduced form by entering household income in logs, which also implies diminishing marginal utility of income. Paid-work and care and household-work hours enter separately and linearly, allowing the value of additional children to be expressed in either type of weekday time. Separate indicators for one and two additional children allow their values to differ across family-size increases, while the interaction with centered log income allows those values to vary with household income. We assume that the unobserved utility terms follow an extreme-value distribution, yielding the conditional-logit choice probabilities.

At household income $Y$, the part-worth of $k$ additional children relative to none is

$$\omega_k(Y) = \alpha_k + \mu k\left(\ln Y - \overline{\ln Y}\right), \qquad k \in \{1, 2\}, \tag{8}$$

with $\omega_0(Y) = 0$. Thus, $\alpha_1$ and $\alpha_2$ are the part-worths of one and two additional children at mean

log income. For a family-size increase from $a$ to $b$, let $\omega_{a\to b}(Y) = \omega_b(Y) - \omega_a(Y)$.

We express these part-worths as compensating variations. For an increase from $a$ to $b$ additional children at household income $Y$, let $CV_{a\to b}(Y)$ solve

$$v(b, Y - CV_{a\to b}(Y), W, H) = v(a, Y, W, H), \tag{9}$$

where $v$ is the systematic component of utility. A positive value is the income a respondent would give up for the larger-family scenario; a negative value is the additional income that scenario would need to provide to leave her indifferent. We solve this condition exactly and report the resulting dollar amount and share of household income at each fielded income level.

The corresponding time equivalents use the paid-work or care and household-work coefficient as the exchange rate. At mean log income, they are

$$TE^{r}_{a\to b} = -\frac{\omega_{a\to b}(\overline{Y})}{\beta_r}, \qquad r \in \{W, H\}, \tag{10}$$

where $\overline{Y} = \exp(\overline{\ln Y})$. A positive value is the number of weekday hours of leisure and sleep the respondent would give up for the larger-family scenario; a negative value is the additional leisure and sleep required to make that scenario equally attractive.

The two education strata were shown different income ranges. Because logit coefficients are identified only relative to the dispersion of unobserved utility, differences in coefficient magnitudes between the strata can reflect differences in choice precision as well as differences in preferences. We obtain the primary valuations by estimating the model separately for each education stratum, so that all coefficients can differ between the two groups. To test whether the two groups share the same preferences once differences in choice precision are allowed for, we also estimate a pooled scaled logit, holding the child and time coefficients and the child-by-income interaction common across strata while allowing the utility scale, the income coefficient and the position effect to differ, and compare the two with a likelihood-ratio (Swait–Louviere) test (Swait and Louviere, 1993). Both analyses use all choice tasks, control for scenario position and cluster standard errors by respondent.

For the first-child comparison, we also estimate the model among respondents with no children, for whom zero to one additional child represents entry into parenthood. Finally, we re-estimate the model with income entering linearly in dollars and with separate coefficients for the four fielded income levels. We retain log income as the primary specification and use these alternatives to assess the role of the money scale.

### 3.2 Hypotheses

#### *3.2.1 Confirmatory hypotheses*

Viewed through the lens of consumer theory, a child is not an ordinary good. It is close to indivisible, largely irreversible, and much of its cost falls on the mother through forgone earnings and time as well as direct expenditure. Our three confirmatory hypotheses ask whether family-size choices

respond systematically to price and income. H1 and H3 test whether demand falls with the money and time prices, while H2 tests whether the value of an additional child rises with household income. All three hypotheses were registered before we examined the main-study data.

- **H1 — Downward-sloping demand in the money price.** As the income sacrifice attached to the larger-family scenario rises, respondents are less likely to choose it: $\theta < 0$ in the reduced-form demand equation.
- **H2 — The value of an additional child rises with household income.** At a fixed proportional money price and time price, the child-by-log-income coefficient is positive: $\mu > 0$. As pre-registered, we report a confidence interval for $\mu$ rather than a one-sided test.
- **H3 — Downward-sloping demand in the time price.** As the leisure and sleep forgone in the larger-family scenario rises, respondents are less likely to choose it: $\phi < 0$ in the reduced-form demand equation.

### *3.2.2 Exploratory hypotheses*

We also test six exploratory hypotheses, H4 to H9, and report two-sided tests and 95 percent confidence intervals for each. H4 was registered as an exploratory hypothesis; the others were not. H7 to H9 examine the shape of demand and rely on attributes randomized within respondents. H4 to H6 instead compare respondents who differ in stated fertility, religiosity or age; because these characteristics are not randomized, the comparisons are descriptive.

- **H4 — Stated ideals are a better guide to choice when women want no more children.** Among childless women choosing between zero and one additional child, those who say they want no children choose in line with that ideal more often than those who say they want one. In a logit of choosing the one-child scenario on an indicator for wanting one child, with intercept $\gamma_0$ and coefficient $\gamma_1$, H4 is $2\gamma_0 + \gamma_1 < 0$.
- **H5 — Stated desired fertility predicts choices.** Respondents who say they want more additional children choose the larger-family scenario more often. We estimate

$$y_{it} = \alpha_{s(i)} + \kappa a_i + \theta P_{it} + \phi T_{it} + \lambda L_{it} + e_{it},$$

  where $a_i$ is the number of additional children respondent $i$ says she wants, her stated ideal family size minus her current children, and $\alpha_{s(i)}$ is an education-stratum fixed effect. H5 is $\kappa > 0$. Respondent fixed effects would absorb $a_i$, so the regression uses stratum fixed effects, with standard errors clustered by respondent.
- **H6 — Baseline demand varies across respondents.** At equal money and time cost, the probability of choosing the larger-family scenario varies with religiosity, desired additional fertility and whether the respondent is aged 30 or over. H6 concerns the coefficients on these three characteristics in a regression of the same form as H5 that includes all three, together with their interactions with the money and time prices. We report the interactions, which show how price

responses vary across respondents, as a supporting analysis.

- **H7 — Demand is lower for a second additional child than for a first.** Let $D_{it}^{g}$ indicate that task $t$ compares zero with one, one with two, or zero with two additional children. We estimate

$$y_{it} = \delta_i + \sum_g \theta_g D_{it}^{g} P_{it} + \eta_{12} D_{it}^{12} + \eta_{02} D_{it}^{02} + \phi T_{it} + \lambda L_{it} + u_{it}, \tag{11}$$

with zero-to-one as the omitted comparison, so that each comparison has its own money-price slope $\theta_g$. H7 is $\eta_{12} < 0$: at equal money and time prices, respondents are less likely to choose two additional children over one than one over none. This is a difference in choice probabilities; the structural estimates give its value in money.

- **H8 — Paid work reduces demand disproportionately at full time.** Let $\phi_4^W$ and $\phi_9^W$ be the effects of four and nine hours of paid work, relative to none, in the specification with a separate coefficient for each level of time use. H8 is

$$\phi_9^W - \tfrac{9}{4}\phi_4^W < 0 :$$

the effect of nine hours is more negative than a linear extrapolation from four.

- **H9 — The care and household-work response is linear.** Care hours take the values 2, 3, 5 and 8, with 2 as the omitted level. Linearity implies

$$\phi_5^H = 3\phi_3^H, \qquad \phi_8^H = 6\phi_3^H,$$

and H9 is the joint test of these restrictions. Together with H8, it distinguishes a cost specific to full-time paid work from a cost that rises with any time taken from leisure and sleep.

## 4 Results

We organize the results by economic question rather than by hypothesis. H1 and H3 appear with the money and time prices, H2 with income, H7 with the value of a first and second additional child, H8 and H9 with the time price, H6 with heterogeneity, and H4 and H5 with stated ideals.

### 4.1 The money and time prices of a child

#### *4.1.1 Demand for a larger family*

Demand for children slopes down in both money and time. Before we impose a functional form, we plot the share of respondents choosing the larger family against the income it costs or provides relative to the alternative (Figure 1). The share declines steadily as the larger family becomes more expensive. The percentage of women choosing the larger-family scenario falls from 69.1 percent when that scenario includes 208 percent more income than the alternative to 24.7 percent when it includes 68 percent less. We plot the fitted linear probability model alongside these shares.

A 10 percent reduction in the larger family's household income lowers the probability of choosing

it by 2.8 percentage points ($\theta = -0.268$, $p < 0.001$). The time price operates in the same direction; the share choosing the larger family falls from 51.7 percent when the larger-family scenario involves nine fewer non-leisure hours than the alternative to 46.6 percent when it involves nine more, and each additional non-leisure hour lowers the probability of choosing it by 0.6 percentage points ($\phi = -0.0063$, $p < 0.001$). The money and time prices are the confirmatory tests of H1 and H3, registered with directional predictions.

Splitting the time price into paid work and care and household work, the two have similar effects on average ($-0.0063$ and $-0.0065$; column 3). However, the linear specification holds each hour type's effect constant across hour levels. Letting it vary reveals that paid and care hours diverge, which we examine next.

#### *4.1.2 The shape of paid-work and care costs (H8–H9)*

Estimating the time price non-parametrically, with a separate coefficient for each fielded hour level, shows that the effect of paid hours is not linear (Table 4). Relative to no paid work, a part-time weekday of four hours raises the probability of choosing the larger family by 1.7 percentage points (standard error 0.9), whereas a full-time weekday of nine hours lowers it by 4.3 percentage points. Moving from part-time to full-time reduces it by 6.0 percentage points (column 2), indicating that women are no less likely to choose the larger-family scenario when it comes with part-time work than with no paid work at all, and that it is the full-time margin, not paid work as such, that deters them.

The care and household effect, by contrast, increases steadily with hours. H8 tests paid-work proportionality directly within the flexible model, $\phi_9^W = (9/4)\phi_4^W$, and rejects it ($p < 0.001$): the full-time effect is more adverse than a linear extrapolation from four hours. H9 is the non-directional joint test that the care effects are proportional to the distance from the omitted two-hour level, $\phi_5^H = 3\phi_3^H$ and $\phi_8^H = 6\phi_3^H$. That test is not rejected ($p = 0.59$), so the care estimates are consistent with linearity over the fielded support; failure to reject is not evidence that linearity holds exactly. Paid and care effects also differ at equal hours ($p < 0.001$). These shape specifications are exploratory; the registered plan enters both hour terms linearly.

A Beckerian model that prices each hour diverted from market work at a constant wage (Becker, 1965) implies a utility cost of paid work that rises in proportion to hours. Our nonparametric estimates depart from this: the cost of paid work is concentrated almost entirely in the transition from part-time to full-time work. Because income and hours are randomized independently, wages play no role in our design, so this concerns the non-pecuniary cost of hours. The pattern is also not the one a fixed cost of labor-force entry would produce (Cogan, 1981): such a cost would deter any paid work relative to none, whereas four hours of paid work carries no penalty. What fits instead may be closer to Goldin's "greedy jobs" (Goldin, 2021): moderate, structured work does not reduce demand for children; it is the transition into an inflexible full-time tier specifically, not hours per se, that is costly — consistent with the value already documented for schedule control (Mas and Pallais, 2017). Our design does not vary job flexibility, and because leisure is the residual, nine-hour

days also leave the least leisure (10 to 13 hours), so we cannot fully separate a full-time penalty from a low-leisure penalty. This also runs counter to a simple story in which women are merely pricing in the part-time wage penalty documented in the labor-market literature (Schrenker, 2023; Backhaus et al., 2023): if that penalty alone drove demand, part-time work should look costly relative to no work at all, yet it does not. This speaks to the child-penalty literature from the opposite direction: that literature documents how children reduce women's paid work (Kleven et al., 2019a; Cortés and Pan, 2023), leaving preferences as the one input to that relationship it cannot observe. Our result adds a different piece: women anticipate the structure of paid work itself, and penalize full-time scenarios whether or not they involve an additional child, which is consistent with an anticipated cost of full-time work that is not specific to childbearing. It also carries into preferences a pattern Goldin (Goldin, 2014) documents in wages: work that rewards continuous, inflexible hours disproportionately over the wage schedule appears to be disproportionately costly in what women want, too, a convexity that realized-earnings data alone cannot attribute to preferences rather than the market. Because we measure only women's demand, this bears on one side of that relationship — the preference realized-outcome data cannot isolate — not on the gender gap itself, which also depends on constraints beyond this design.

The aversion to full-time work is not confined to choices about children. Even when neither scenario includes an additional child, and the two differ only in household income and the respondent's hours, women are 9.6 percentage points less likely to choose nine hours of paid work over four (SE 1.7). The full-time penalty therefore reflects how women value full-time work itself, not only how they value it alongside an additional child.

#### *4.1.3 The value of an hour*

The price sensitivity of demand to time, relative to its sensitivity to money, implies a value of an hour of the respondent's own time: a woman would need 2.35 percent more annual household income to accept one additional weekday hour of paid work or care, taken from leisure and sleep, and would give up as much to avoid one (SE 0.43; 95 percent interval 1.51 to 3.19). Evaluated at the median income within each education group, $95,000 and $175,000, this is $8.59 an hour for non-college women and $15.82 for college women.

This is a value of time in the sense used in the stated-preference literature on work arrangements (e.g., Mas and Pallais, 2017), not a wage: household rather than personal income is the relevant resource here, and the estimate averages paid work and care, which Section 4.1.2 shows behave differently. Because the time response strengthens across tasks (Appendix H), we read this value as conservative.

This direct elicitation is new for the fertility literature as well as the gender one. The modern economics of fertility places the time cost of children at the center of its account of low fertility (Doepke et al., 2023), but the evidence for that cost is indirect, coming from variation in childcare prices, availability, or parental leave — for instance, childcare expansions have been found to raise completed fertility where cash transfers have only temporary effects (Bergsvik et al., 2021). None of

this evidence prices a woman's own time directly, or puts it on the same footing as money. Because our design elicits both in the same choice, the value of an hour above is directly comparable to the value of a dollar, in a way that indirect policy evidence cannot offer.

### 4.2 The value of a first and a second additional child (H7)

#### *4.2.1 Choice probabilities*

The probability of choosing the larger family falls sharply from the first additional child to the second. Across all tasks, respondents choose one additional child over none in 52.0 percent of tasks, two over one in 42.9 percent, and two over none in 46.8 percent.[4]

The H7 regression in Equation (11) adjusts for the money and time prices in each task and compares each respondent's choices only with her own. At equal prices, the probability of choosing two additional children over one is 9.8 percentage points lower than the probability of choosing one over none ($\hat{\eta}_{12} = -0.098$, SE 0.011), close to the 9.1-point gap in the raw shares. We reject the null hypothesis of no difference ($p < 0.001$, two-sided), and the estimate has the negative sign that H7 predicts. The probability of choosing two over none is 5.3 points lower than one over none ($\hat{\eta}_{02} = -0.053$, SE 0.010). The gap holds across the full range of fielded money prices (Appendix Figure A2). Section 4.5 shows that the low value of a second additional child persists even among women who say they want two more.

#### *4.2.2 Willingness to pay*

The willingness-to-pay estimates give the size of this decline in money (Table 5). Evaluated at the median household income in each education group, college women would give up $20,162 in annual household income to move from zero to one additional child. The corresponding estimate for non-college women is $4,643, although it is not statistically distinguishable from zero. Willingness to pay for the second child is strongly negative: non-college women would require approximately $26,300 in compensation to move from one to two additional children, and college women approximately $29,300.

#### *4.2.3 By current parity*

These estimates average across women at different current parities. For childless women, moving from zero to one additional child represents entry into parenthood; for mothers, the same comparison represents progression to a later birth. Preferences may differ by current parity for several reasons. Mothers may be positively selected on their desire for children. Experience of parenthood may also change a woman's assessment of its costs and benefits: women appear to underestimate the employment costs of motherhood before a first birth (Kuziemko et al., 2018). Mothers may also

[4]This can also be seen in the 365 tasks in which the two scenarios offer identical household income, paid-work hours and care hours, the corresponding shares are 54.3, 32.5 and 50.4 percent. These shares are less precise, each with a margin of error of about nine percentage points.

have already incurred some of the fixed costs of having children, but may be closer to their desired family size.

Table 6 shows that the valuation of the first additional child depends strongly on whether a woman is already a mother. Childless non-college women would require approximately $16,200 in compensation to enter parenthood. The corresponding estimate for childless college women is approximately $11,800, although it is not statistically distinguishable from zero. Mothers, by contrast, would give up approximately $22,400 among non-college women and $54,300 among college women to have one additional child. Positive willingness to pay for the first additional child is therefore concentrated among women who are already mothers.

The fall in value from the first to the second additional child, however, is common to both groups. Childless women would require approximately $47,700 among non-college women and $44,400 among college women to move from one to two additional children, considerably more than they require for the first. Among mothers, the corresponding compensation is approximately $6,900 and $8,800, and neither is statistically distinguishable from zero. Mothers who would pay for a first additional child therefore place no detectable value on a second. The decline in value from the first to the second additional child persists within both parity groups, rather than being driven by averaging across childless women and mothers.

### 4.3 Income and the valuation of additional children

#### *4.3.1 The registered test*

Section 4.1 establishes how respondents' choices respond to changes in the money and time prices presented in the scenarios. We next examine how the valuation of an additional child varies with the income level of the scenarios shown to respondents. We test this question using $\mu$, the child-by-log-income interaction in the reduced-form regression reported in column (2) of Table 3. Because income enters in logs, this coefficient captures how the response to an additional child changes when the income levels in both scenarios rise proportionally, leaving their relative income unchanged. We report the 95 percent confidence interval for this estimate, as pre-registered: [0.012, 0.045]. The point estimate, $\mu = 0.028$, implies that a 10 percent increase in both scenario income levels raises the probability of choosing the larger-family scenario by 0.27 percentage points per additional child. By comparison, increasing the larger-family scenario's income alone by 10 percent raises that probability by about 2.3 percentage points through the price channel alone. Raising one scenario's income also moves the child-by-income interaction, which adds between 0.3 and 0.5 percentage points depending on the family size on offer, so this figure is illustrative rather than a complete marginal effect.[5]

The structural model, which is estimated using all choice tasks rather than only those in which the scenarios differ in family size, gives a very similar result (Appendix Table A3). For a one-child

[5]This calculation uses the income coefficient in column (2) of Table 3, which includes the child-by-log-income interaction. The 2.8-percentage-point H1 estimate reported above uses the coefficient in column (1), from the specification without that interaction.

difference, a 10 percent increase in both scenario income levels raises the predicted probability of choosing the larger-family scenario by 0.25 percentage points, compared with 0.27 percentage points in the reduced-form regression.

The income interaction provides experimental evidence on the longstanding question of how demand for children varies with income. The natural benchmark is homothetic preferences over consumption and children: the value of an additional child scales in proportion to income. In our reduced form, this restriction corresponds to $\mu = 0$. The confidence interval lies above zero, implying that the value of an additional child instead rises more than proportionally with scenario income. The result therefore represents a modest rejection of homotheticity, consistent with children being a luxury good in this setting. This matches quasi-experimental evidence that children behave as a normal, income-elastic good (Black et al., 2013), and is consistent with the marketization channel proposed to explain why higher-income parents now have more of them (Bar et al., 2018); our randomized income variation identifies the same pattern without relying on realized income shocks or cross-sectional comparisons. A 10 percent rise in both incomes raises the probability of choosing the larger family by only 0.25 percentage points, against a 19.0-point gradient by religiosity (Table 7). A continuing shift away from pro-fertility norms, documented in rich countries including the United States (Inglehart, 2021) and part of the "shifting priorities" that Kearney and Levine (2025) emphasize, could therefore outweigh a mild luxury-good effect of income growth.[6]

Figure 2 shows how the structural valuations vary with household income. It reports willingness to pay for the larger-family scenario as a share of household income, separately for the zero-to-one, one-to-two and zero-to-two comparisons, at each income level shown to each education group.

Across all three comparisons and in both education groups, willingness to pay rises as a share of income. However, this does not overturn the negative valuation of the second additional child: willingness to pay in the one-to-two comparison remains negative at every income level. In the zero-to-two comparison, non-college women require compensation at every income level, whereas the college estimates are close to zero and not statistically distinguishable from it. Because the two groups were shown different income ranges, this comparison combines differences by education with differences in income. Where the ranges overlap, the gap persists: non-college women require 17.6 and 14.8 percent of income at \$140,000 and \$185,000, compared with 6.8 and 3.9 percent for college women at \$130,000 and \$175,000.

#### *4.3.2 Slutsky decomposition*

The money-price response can be split into a substitution effect and an income effect, the same decomposition labor economists apply to a wage change. Because compensated responses in discrete choice are specific to the direction of the price change and to the family-size contrast, this split is a

[6] From 2007 to 2023–24 the share of US adults saying religion is very important fell by 14–18 points, from 56 to 38–42 percent (Pew Research Center, 2025). Treating our unconditional religiosity gradient (22.2 points, or 7.4 per step on the four-point scale) as the effect of changing norms, even the smallest move (one step for each leaver) lowers demand by 1.0–1.3 points. Real median household income rose 15 percent (U.S. Census Bureau, 2025), which adds 0.4 points at 0.25 points per 10 percent.

set of contrast-specific coefficients rather than a single number; Appendix C gives the derivation and the full table. Across the zero-to-one, one-to-two, and zero-to-two contrasts, substitution accounts for approximately 81 to 90 percent of the money-price response, and the income effect for the remaining 10 to 19 percent. The money-price response to an additional child is therefore mostly a genuine relative-price effect rather than households choosing fewer children because the higher price has left them poorer.

### 4.4 Heterogeneity

Our estimates reveal substantial heterogeneity in the level of fertility demand, especially at the first-additional-child margin and at entry into parenthood. However, the curvature of demand is remarkably consistent: although groups differ in how they value both the first and the second additional child, the second is a much less valued margin across education, parity, religiosity, desired additional fertility, and age.

We distinguish heterogeneity in baseline demand from heterogeneity in price responses. The baseline choice probability indicates how likely a group is to choose the larger family when the measured costs are equal, while willingness to pay gives the compensating change in income that offsets the estimated utility gain or loss from an additional child. The price response instead measures how the probability of choosing the larger family changes as its money price increases. A group can therefore place a higher value on an additional child without exhibiting a stronger price response. Empirically, heterogeneity is more pronounced in baseline demand and valuations than in money-price responsiveness.

The average estimates conceal substantial variation across respondents. Across the unequal-child tasks, 13.6 percent of women never chose the larger-family scenario, 10.0 percent always chose it, and 76.4 percent chose the larger family in some tasks but not in others. When income and time are held equal, 54.3 percent choose one additional child over none. The one-versus-two choice share provides a model-free counterpart to the negative average valuation: when resources are held equal, respondents choose one additional child over two more often than the reverse. Even so, 32.5 percent choose two, indicating that the second additional child remains attractive to a substantial minority.

The parity estimates reported in Table 6 identify one source of this variation. Differences between childless women and mothers are concentrated at the first-additional-child margin, while the decline in value from the first to the second remains within both groups.

We next compare the college and non-college estimates. This comparison requires accounting for differences in choice precision because conditional-logit coefficients are identified only relative to the variance of unobserved utility. Differences in coefficient magnitudes may therefore reflect differences in the consistency of choices as well as differences in preferences.

We use a pooled scaled logit to assess whether the relative preference parameters differ once this scale difference is taken into account. The estimated utility scale for college women relative to non-college women is 1.17 (SE 0.14). A likelihood-ratio test does not reject equality of the child part-worths, time coefficients, and child-by-income interaction, while allowing the utility scale,

income coefficient, and position effect to differ between strata ($\mathrm{LR}(4) = 7.08$, $p = 0.132$). The data therefore provide no clear evidence that the two groups have different relative preference structures.

At the median income of each stratum, estimated willingness to pay for the first additional child is higher among college women. This may partly be because college respondents were shown higher incomes: the preceding results show that the valuation of children rises modestly with income, so the two estimates are evaluated at different points along this gradient. The compensation required for the second additional child is similar across the two groups. More generally, both display the same ordering across margins, with positive willingness to pay for the first additional child and negative willingness to pay for the second. Together with the scaled-logit test, these estimates provide little evidence of a pronounced difference in the underlying structure of preferences by education.

We also explore heterogeneity along three further dimensions: religiosity, desired additional fertility, and age. Table 7 reports reduced-form estimates of each characteristic's association with baseline demand and of its interactions with the randomized money and time prices. Columns 1–3 enter the characteristics one at a time and column 4 enters them together. Respondent fixed effects absorb characteristics that do not vary across a respondent's tasks, so the baseline-demand coefficients come from a separate regression with education-stratum fixed effects. We adjust for multiple testing using the Romano–Wolf procedure (Romano and Wolf, 2005; Clarke et al., 2020), treating each column as one family: a characteristic's three coefficients in columns 1–3 and all nine coefficients in column 4. Each bootstrap replication re-estimates both regressions on the same resampled respondents, so the adjustment accounts for the correlation between coefficients estimated in different regressions.

We measure religiosity by the importance of religion in a respondent's life rather than religious affiliation because the intensity of religious commitment is more directly related to fertility intentions and behavior than nominal affiliation (Hayford and Morgan, 2008). Desired additional fertility, defined as stated ideal family size minus current parity, captures whether a respondent is already satiated at her stated ideal. Including it alongside religiosity and age helps distinguish their associations with demand from differences in how many respondents in each group already have the number of children they say they want. We also divide the sample between women under 30 and those aged 30 or over. We choose this cutoff because the scenarios describe family size ten years in the future: this is more likely to approximate completed fertility for women in the older group, whereas younger women may plan births beyond the period described in the experiment.

The baseline-demand rows of Table 7 show how these characteristics predict the propensity to choose the larger-family scenario when its measured money and time costs are held equal. In the joint model in column 4, moving from "not at all important" to "very important" on the four-point religiosity scale is associated with a 19.0 percentage-point increase in this probability, each additional desired child with a 7.9-point increase, and being 30 or older with a 7.0-point increase. Entering each characteristic alone, in columns 1–3, gives 22.2 (from the lowest to the highest religiosity), 8.1 (desired additional fertility per child), and 3.6 points (age 30 and over).

The interaction rows show how the responses to the money and time prices vary with these

characteristics. The adjusted tests yield two findings. First, the negative money-price response becomes stronger with each additional child desired: women who want more children are more likely to choose the larger family at a given price, but they also respond more strongly when its relative income falls. Second, the negative time-price response weakens across the religiosity scale. An additional hour reduces demand among women for whom religion is not at all important, but has almost no effect among those for whom it is very important. The remaining interactions do not survive the multiplicity adjustment: religiosity does not clearly predict the money-price response, desired additional fertility does not clearly predict the time-price response, and neither response varies detectably with age.

Table 8 reports willingness to pay from models estimated separately within each subgroup. This allows comparisons of valuations without conflating them with differences in how consistently the groups make their choices: willingness to pay compares the child and income coefficients estimated within the same model, so the subgroup-specific logit scale affects both equally and cancels from the comparison. The common ordering of the child margins is also not imposed by the model. The utility values of one and two additional children are estimated separately, with no restriction that the second incremental value must be lower than the first. The negative second-child estimates are therefore a feature of the choices rather than a mechanical consequence of the specification. Second-additional-child willingness to pay is negative in every subgroup, and 11 of the 12 confidence intervals exclude zero; the exception is college women for whom religion is important, whose point estimate remains negative. At the same time, the estimates reveal large differences in the valuation of the first additional child. Among college women, willingness to pay changes from approximately −$50,000 among those who want no more children to $51,000 among those who want at least one more. Religiosity produces a similarly pronounced change in both education strata, although the negative estimate for less-religious college women is not statistically distinguishable from zero. Differences by age are less systematic and less precisely estimated, possibly because the structural analysis divides the sample by both age and education.[7] Across the dimensions examined, women differ sharply in their valuation of the first additional child, but no subgroup reverses the lower valuation of the second.

### 4.5 Stated ideals and contextualized demand

Ideal family size and fertility intentions have been used for decades to measure fertility preferences in demographic surveys. We therefore asked respondents how many children they would ideally like to have in total, allowing us to compare this conventional measure with the choices they made in the experiment. Their responses reproduce the familiar two-child ideal: the mean is 1.83 children, and the mode is two, reported by 33.7 percent of respondents.

Our experiment elicits fertility preferences in a more specific setting. Respondents choose between alternative lives ten years in the future, each of which specifies that they are living with a

[7] Age and birth cohort are collinear in a single cross-section, so this comparison cannot separate a life-cycle effect from a cohort difference in preferences that happens to track age at the time of the survey.

partner, their number of additional children, their household income, and the allocation of their own time.

Translating respondents' ideal family size into the number of additional children they say they want partially explains the low valuation of the second additional child. After subtracting current children from the stated ideal, 46.4 percent of respondents want no more children, 21.8 percent want one more, and 31.9 percent want two or more. Thus, more than two-thirds of respondents do not say that they want two additional children.

However, the low valuation of the second additional child is not confined to respondents who say they want fewer than two additional children. Desired additional fertility predicts choices in the experiment: each additional child a respondent says she wants is associated with a 7.9 percentage-point increase in the probability of choosing the larger family (SE 0.6 percentage points). However, while valuations generally rise with desired additional fertility, the estimated value of a second additional child remains low among women who say they want two more.

The blue and pink bars in the upper panel of Figure 3 show willingness to pay for the first and second additional child, by desired additional fertility. Among women who want two more, the value of the second additional child is −$13,400 a year (95 percent confidence interval −$22,600 to −$4,200).

The lower panel restricts the sample to women aged 30 or over. The scenarios describe life ten years from now, which for these women covers nearly all of their remaining childbearing years, so the number of additional children they choose is close to a choice of completed family size. Valuations are generally higher among women aged 30 or over. Among those who want two more, however, the value of the second additional child is imprecisely estimated (−$3,500 a year, 95 percent confidence interval −$19,200 to $12,200).

Each scenario states household income and the respondent's weekday hours, and these vary independently of the number of children, so willingness to pay values a child who costs the mother no earnings and no extra time. In practice, mothers' earnings fall after a birth and their hours of paid and unpaid work rise. We call the value of this lost income and time the child penalty, extending the usual earnings penalty to include the mother's time. The figure also shows valuations net of this penalty: the purple and orange bars subtract low and high estimates of it from willingness to pay (Appendix E). We do not subtract the cost of goods such as food, clothing and childcare: respondents know these must be paid for out of the household income stated in the scenario, so they are already reflected in their choices. Net of the child penalty, the second additional child is valued lower still. Among women who want two more, it requires compensation in both samples, while the first additional child keeps a positive value among women who want more children.

The two measures need not coincide because the ideal-family-size question leaves the circumstances of the choice open. Respondents may report the family size they would prefer under conditions they have in mind, whereas the experiment elicits valuations across a series of specified futures. An ideal of two may therefore apply only under a narrower set of circumstances and need not imply a general preference for a second child.

The low value of the second child among women who want two more is part of a broader asymmetry: a stated desire for no more children is a much more reliable guide to choice than a stated desire for more. Registered H4 (Section 3.2) isolates this asymmetry in the simplest comparison the experiment allows, among childless women choosing between no children and one child. For these women, the number of children they want in total is also the number they want in addition. Women who say they want no children choose none in 82.0 percent of these tasks, whereas women who say they want one child choose one in only 49.7 percent (Table A8; registered log-odds contrast $-1.531$, SE 0.162, $p < 0.001$). Research on fertility intentions finds the same asymmetry in behavior: women who intend to have no more children are more likely to realize that intention than women who intend to have another (Schoen et al., 1999).

Respondents reported their ideal family size after the ten choice tasks, so there is a possibility that exposure to the choice tasks may have influenced their answers. However, we do not find evidence of this. Our stated ideals are close to those in the NSFG once we standardize for differences in composition between the two samples (Section 5.1). We also test whether the tasks a respondent was shown affected her stated ideal. As the scenarios were randomly assigned, we regress the stated ideal on the mean number of children shown across her twenty scenarios and the mean money and time prices of the larger-family scenario, controlling for education, which determines the income levels she was shown. None of the three coefficients is statistically significant, and they are jointly insignificant ($p = 0.70$). The same holds for desired additional fertility ($p = 0.57$).

### 4.6 Robustness

Table 9 reports the confirmatory estimates under the registered alternative weighting and analysis samples. The money- and time-price coefficients remain negative and the child-by-income interaction remains positive throughout, leaving all three confirmatory conclusions unchanged. Post-stratification weighting has almost no effect because the realized age-by-education composition is already close to the population composition used to construct the weights.

The most informative variation concerns how respondents engaged with the choice tasks. Among the slower two-thirds of respondents, the time-price coefficient increases in magnitude from $-0.00629$ in the primary sample to between $-0.00879$ and $-0.00930$, depending on whether response time is measured over the choice tasks alone or together with the instructions. The money-price response and child-by-income interaction are also larger in these samples. The fact that all three coefficients move away from zero is consistent with less precise choices among the fastest respondents attenuating the estimated trade-offs.

The response to time also becomes substantially stronger as respondents progress through the experiment, increasing from $-0.00331$ in the first five tasks to $-0.00920$ in the final five. The money-price response increases by considerably less. This differential change is consistent with respondents becoming more familiar with the time-use profiles, which require them to consider paid work, care and household work, and the time remaining for leisure and sleep. Both engagement comparisons therefore produce a larger time response than the primary estimate, which can be read

as a conservative measure relative to the estimates among slower respondents and in the later tasks.

A related concern is that respondents who consistently select higher income may be attending only to that attribute. Among the 1,995 respondents who faced at least one task in which the scenarios differed in income, 351, or 17.6 percent, always selected the higher-income scenario. However, these respondents remain strongly responsive to time when the two scenarios offer the same income. An additional hour of paid work or care and household work reduces their probability of selecting a scenario by 3.31 percentage points, compared with 1.60 percentage points among other respondents. Consistently selecting higher income when it differs therefore does not appear to reflect non-attendance to the time attribute.

We next examine whether the structural valuations depend on how the income response is specified. At the median income of each education stratum, replacing log income with the registered linear-dollar scale changes the estimated income equivalents by no more than approximately $4,800. Allowing each fielded income level to enter separately also preserves the signs and ordering of the valuations, although the income response flattens at the top of the college income range.

The primary conditional-logit model represents respondents within each education stratum using common valuations of one and two additional children. If these valuations differ substantially across respondents, imposing common coefficients in a nonlinear choice model could affect the mean valuations rather than merely conceal their dispersion. We therefore estimate mixed-logit models in which the valuations of one and two additional children, and the response to income, vary across respondents and may be correlated. The median valuations are close to the fixed-logit estimates and slightly larger in magnitude, and the ordering of the first- and second-child valuations is unchanged.

The secondary results exhibit the same stability across the alternative samples. In particular, religiosity, desired additional fertility, and age 30 or older remain positively associated with baseline demand for children.

The robustness analysis therefore adds one substantive qualification to the primary results. We retain the full-sample time-price estimate as the primary estimate, but interpret it as conservative because both engagement analyses produce stronger responses to time. No comparable qualification emerges for the money-price response, the income gradient, or the ordering of the first- and second-child valuations. Appendix H reports the full results and additional checks.

### 4.7 Interpreting the estimated attribute effects

The case for eliciting preferences through hypothetical scenarios rests on holding respondents' beliefs fixed by giving them as close to full information as possible about the context (Almås et al., 2024, p. 957). Because scenario attributes are randomized, respondent characteristics cannot generate conventional omitted-variable bias. However, a displayed attribute may change respondents' beliefs about an unshown feature of the scenario. The estimated coefficient therefore captures the effect of the displayed attribute on stated choice, including any inferences it generates. This is the *information-equivalence* concern of Dafoe et al. (2018), and Adrjan et al. (2026) demonstrate it experimentally in a job-choice setting: randomly varying advertised amenities changes beliefs about

unlisted job characteristics, and posted wages themselves signal features such as workplace culture and stress. We find little sign that such inferences drive our estimates: the partnered comparison limits partner inference to about 3 percent of the money-price coefficient, and job-stress inference is bounded at 5.5 percent.

Unshown features can enter respondents' choices in three ways. The first is *causal consequences*: changes respondents expect the displayed attribute itself to produce, such as the expense and fatigue of an additional child, or the housing and security that higher income buys. These belong in the valuation of the displayed future. The second is *household allocation assumptions*: the scenarios do not say whether household income comes from the woman's earnings or her partner's, or who supplies the care she does not. These matter when the estimates are compared with a setting whose allocation differs from what respondents assumed (for example, a cash transfer raises household income without changing who earns it), but they do not change what the estimates measure within the experiment. The third is *inferences about associated attributes*: a respondent may read into the displayed income a more demanding job, or a partner with particular characteristics, that the scenario neither states nor implies. It is this third channel that can make the money-price coefficient differ from the value of the price itself.

We do not assume a unitary household, nor that the relationship is permanent. Rather than imposing a single household utility function, we leave the allocation of income and care between partners unspecified, and the evidence below examines whether respondents' assumptions about that allocation, and the partners they infer, bias the money-price coefficient. Respondents may also anticipate that the relationship could end, and value earnings of their own as protection for themselves and their children; the open-text feedback indicates that some did. The valuations should therefore be read as incorporating the possibility of separation, which is also present when women make actual fertility decisions. The design does not model bargaining between partners directly; Doepke and Kindermann (2019) provide the theoretical framework for that question. Asking both partners to price the same prospective child would bring bargaining into the design, and we return to this in the Conclusion.

Following Adrjan et al. (2026, Appendix D), suppose a respondent's imputed value $\hat{A}$ of an unshown attribute moves with the money price $P$. The estimated money-price coefficient is then

$$\hat{\theta} \;=\; \theta_{\text{pure}} \;+\; \beta\,\frac{\partial \hat{A}}{\partial P}, \tag{12}$$

where $\theta_{\text{pure}}$ is the effect of the price alone, $\beta$ is the weight the choice places on the imputed attribute, and $\partial\hat{A}/\partial P$ is how far the imputed attribute moves with the price. The direction of the difference depends on the attribute. If higher income signals a partner with more valued characteristics, the lower-income option also looks worse on that unshown dimension, and $\hat{\theta}$ lies further from zero than $\theta_{\text{pure}}$. If higher income signals a more demanding or stressful job, the higher-income option looks worse, and $\hat{\theta}$ lies closer to zero.

Adrjan et al. (2026) derive this decomposition for a job-choice experiment and describe it as

"precisely the structure of omitted-variable bias,"[8] with beliefs about unlisted attributes in the role of omitted regressors. They measure how listing one attribute shifts beliefs about others, using a belief-elicitation module, but not how much respondents value the attributes inferred, and so do not quantify the resulting bias. Our survey has no belief-elicitation module, so we cannot measure either term; we instead examine whether respondents' own accounts and the structure of the sample show signs of these inferences.

**Open-text feedback.** Of the 1,995 respondents in the primary sample, 1,369 left nonblank feedback. Coding these comments with deliberately inclusive definitions, 42 respondents mentioned a household allocation assumption or an associated-attribute inference: 24 referred to the source or control of household income, 16 to the allocation of household work and care, and seven to the job implied by the displayed income and paid-work hours. None inferred anything about a partner from household income beyond how much they work or earn. A further 34 respondents weighed whether the displayed income could support the number of children shown; we classify these as causal consequences of family size, the first channel above, not as inferences. Comments are unprompted and will understate how often respondents made such inferences; they show which considerations were salient rather than how common they were. The seven job-related comments concerned the wage or occupation implied by the income and hours shown; none mentioned workplace culture or stress, the channels Adrjan et al. (2026) document.

**Partnered respondents.** Respondents in a relationship indicated, in the cognitive interviews, the pilot and the main study, that they generally pictured their existing partner across scenarios, whatever the displayed income. Unpartnered respondents must imagine a partner, and may infer a more valued one from a higher income. If that inference contributed materially to $\hat{\theta}$, their money-price coefficient should be more negative than that of partnered respondents. It is not: $\hat{\theta} = -0.279$ among the 1,265 respondents in a committed relationship and $-0.251$ among the 727 who are not, a difference of 0.028 (standard error 0.019) in the opposite direction. The confidence interval excludes a difference of more than 0.009 in the direction partner inference would produce, about 3 percent of the partnered coefficient. The time-price coefficients are $-0.0060$ and $-0.0068$, and the two prices do not differ jointly between the groups ($p = 0.31$). The child-by-income coefficient is 0.035 among partnered and 0.017 among unpartnered respondents ($p = 0.31$ for the difference). Partnered and unpartnered women differ in other ways, so this comparison is consistent with a small role for partner inference rather than proof of one.

**Job quality.** Respondents may also infer something about the job behind a given income. Adrjan et al. (2026) find that workers shown a higher posted wage expect better amenities, such as a

[8] Sensitivity tools for omitted-variable bias, such as the robustness value of Cinelli and Hazlett (2020) or the coefficient-stability approach of Oster (2019), do not apply here. They bound confounding from a pre-existing trait correlated with both treatment and outcome, whereas the inferred belief is a consequence of the randomized attribute itself: a mediator, not a confounder.

pension or team events. Such inferences would make a higher income more attractive than the money alone, biasing $\hat{\theta}$ away from zero and making our willingness-to-pay estimates conservative. The exception is stress: workers also expect a better-paid job to be more stressful, which would bias $\hat{\theta}$ toward zero. External estimates of the value workers place on a less stressful job suggest that this bias is small. Maestas et al. (2023) value a move from a fast-paced to a relaxed job at 4.4 percent of the wage, so even if every respondent pictured a relaxed job at the lowest income in the design and a fast-paced one at the highest, $\hat{\theta}$ would understate the effect of the price alone by at most 5.5 percent (Appendix I). The actual bias is likely to be much smaller, since there is little indication that respondents made this inference: none of the job-related comments mentioned stress.

**Pace of choice.** Respondents spent a median of about 11 seconds on each comparison. At that pace they could readily apply familiar assumptions about earning and caring roles, but had less opportunity to work through the implications of several unshown attributes for each scenario.

Together, these results suggest that inferences about associated attributes were not prominent in respondents' reasoning. For the channel the data can size, partner inference from household income, the partnered comparison limits any effect on the money-price coefficient to about 3 percent. The job-stress channel would push the coefficient toward zero while partner inference would push it away, so the two bounded channels work in opposite directions.

## 5 Comparing our estimates to external evidence

We compare our estimates to independent evidence in three ways: stated fertility ideals against a national survey, each of our three confirmatory hypotheses against natural experiments and revealed-preference evidence from other settings, and the exploratory heterogeneity in demand by religiosity against demographic survey data. Throughout, we make explicit the differences in population, intervention, outcome and time horizon between our design and each external source, rather than deferring them to a concluding discussion.

### 5.1 Stated ideals and external survey evidence

Our stated-ideal measure — the share of respondents whose ideal family size exceeds their current parity — compares closely to external survey data; this is a different comparison from the one already drawn between stated ideals and contextualized demand earlier in the paper. Benchmarking against the 2022–2023 NSFG (RWANT, which asks whether a respondent wants a or another child and so preserves the personal-desire construct, unlike the general-family-size measures in the GSS or Gallup), and restricting both samples to women aged 23–35: 53.6% of our respondents report an ideal above their current parity, against a comparable NSFG share of 57.4% (SE 2.0) once standardized to our joint distribution of parity and college attendance — a gap of −3.8 points (SE 2.3) that narrows to −2.9 points (SE 2.4) under a richer regression standardization. Some of this gap likely reflects timing rather than a genuine difference: the comparable NSFG share fell from 67.8% in 2017–2019 to 57.4% in 2022–2023, and extrapolating that trend to our 2026 field date

places the benchmark at 49.4–54.7%, comfortably containing our 53.6%. A secondary comparison of the full distribution of numeric ideals against NSFG parity plus expected additional births is similarly close (both modal at two children, means of 1.83 and 1.71, total variation distance 0.040). The appendix details the standardization and extrapolation.

### 5.2 The confirmatory hypotheses

We compare the estimates behind our three confirmatory hypotheses with the closest available external evidence. Two limits apply throughout: external studies recover realized behavior rather than the stated preference our design measures, and comparing magnitudes requires converting each study's income or time shock into terms comparable to ours, which we make explicit for each case below.

H1 predicts how income changes shift demand for one additional child; we benchmark $\hat{\theta} = -0.2682$ against cash transfers of comparable proportional size, ordered here from the longest external observation window to the shortest, since even the longest falls far short of our ten-year measure. González (2013) and González and Trommlerová (2023) track Spain's now-canceled €2,500 one-off benefit — 0.77% of household income once annuitized to the same permanent-income terms as our design, implying a predicted 0.21pp — over the roughly 30 months it was continuously in force, recovering 0.24 and 0.12pp respectively — the latter a re-estimate, not an independent replication. Family 500+ is a per-child transfer, so its recipients already have at least one child; benchmarking it against the pooled $\hat{\theta}$ therefore mixes in childless women who cannot receive it. Restricting the money-price estimate to mothers gives $\hat{\theta}_{\text{mothers}} = -0.2721$ ($SE = 0.0135$), and a predicted response of 2.6pp.[9] Against this, Bokun (2024) recovers 1.5 of a predicted 2.6, and Gromadzki (2024) identifies only the immediate, roughly nine-month response to the same transfer and recovers just 0.22 of the same 2.6. The Polish pair, separate studies of the same transfer, again shows shorter windows recovering less (0.22 vs. 1.5 of the predicted 2.6pp), consistent with demand accumulating over several years rather than emerging immediately. The appendix details each conversion, including the assumed baseline fertility rate needed to convert Gromadzki's and González's aggregate national counts into individual probabilities, the largest source of uncertainty here.

H2 measures how demand shifts with overall scenario income, independent of any child-contingent price; we benchmark $\hat{\mu} = 0.028$ against unconditional wealth shocks. Cesarini et al. (2023) use randomized lottery-prize size among Swedish players to estimate wealth effects on fertility by sex; the female estimate is the relevant one for our all-women sample. A $150,000 win increases the number of children born to women within ten years by 0.0078 ($SE = 0.0189$). Converting the prize to a permanent annuity, following their own convention (20 years, 2% discount rate), a $100,000 prize equals a 22.4% increase in median disposable income in their sample; applying $\hat{\mu}$ to a shock this size implies 0.0057 expected additional children (95% CI $[0.0024, 0.0091]$), against their rescaled estimate of 0.0052 ($SE = 0.0126$). Our predicted interval lies entirely inside theirs, though the

[9] This estimate is statistically indistinguishable from the pooled coefficient of −0.2682, which predicts the same 2.6pp.

comparison rests on one study and an individual rather than household income base. Since 38% of winners were married, revaluing their share of the sample at roughly double the individual-income base lowers the implied shock to about 16% and the prediction to 0.0043 children, just below the actual 0.0052 (appendix).

For H3, the closest external evidence comes from extensions of paid parental leave (Dahl et al., 2016; Andersson et al., 2006), but these provide too little time to be comparable: leave frees a few months around a single birth, whereas our scenarios fix a woman's daily time use for ten years.

### 5.3 Religiosity: an exploratory comparison

Heterogeneity in baseline demand provides a further check on external validity, one that does not depend on a natural experiment. If stated choices reflect the preferences that govern actual fertility, characteristics that predict demand in the experiment should predict realized fertility in the same direction. Religiosity is the natural case: it is a stable characteristic, the NSFG measures it comparably, and its association with fertility is well documented. In our data, moving from "not at all important" to "very important" on the four-point scale is associated with a 19.0 percentage-point increase in the probability of choosing the larger family (Table 7). Using the 2002 NSFG, Hayford and Morgan (2008) find that women for whom religion is very important have a total fertility rate half a child higher than those for whom it is not important. Applying their method to the 2022–2023 NSFG, we find that this gap has persisted: at 0.70 children, it is at least as large as in 2002, although fertility is lower in every group. Because period rates are also sensitive to the timing of births, we examine completed fertility among women aged 40–49, for whom the gap is 0.75 children (Appendix Table A14). The experiment therefore recovers a gradient that is large and persistent in realized fertility, whether it reflects pronatalist norms or church-provided services that lower the cost of childrearing (Berman et al., 2018). Respondents' choices thus reflect considerations behind actual fertility decisions, not artifacts of the hypothetical setting.

## 6 Conclusion

We designed and fielded a discrete choice experiment to put a price, in both money and time, on an additional child: something the existing literature on fertility has so far only inferred indirectly, through a structural model fit to realized behavior. Demand for a larger family slopes down in each price, as pre-registered, and rises with household income by more than a homothetic model predicts, consistent with a luxury good rather than a necessity whose value simply tracks resources.

The paper's sharpest result is that this price falls a long way between the first and the second additional child, consistent with recent evidence that parity progression, not childlessness, accounts for most of the fertility decline in advanced economies (Geruso and Spears, 2025). At equal income and time, mothers give something up for a first additional child, but childless women require compensation even for a first. On average, however, the estimated value of a second additional child is negative in each of these groups: college and non-college women, childless women and mothers, and even those who say they want two more children of their own. That last fact bears on how the

persistence of a stated two-child ideal (Hagewen and Morgan, 2005; Sobotka and Beaujouan, 2014) should be read: much of the argument that fertility has fallen below what people want rests on that stated ideal, taken as a preference that survives once its price is fixed. Our results suggest it is considerably softer than that: contextualized against an explicit cost, demand for a second additional child is weak even among the women who name it as their goal.

The shape of the time price adds a second point. Care and household time costs demand roughly linearly, but paid work does not: the cost of a woman's own time is concentrated almost entirely in the transition from part-time to full-time work, not in hours as such.

Our discrete choice experiment on fertility opens onto a research agenda well beyond this paper. Because both partners can answer the same choice tasks, it extends directly to eliciting each side's price for the same additional child and asking when, and for whom, those prices diverge; this is a question at the heart of bargaining models of fertility that has so far had no comparable empirical handle. Because money and time separate cleanly here, the same design travels to countries whose family policies differ sharply in exactly this mix, turning a hypothetical trade-off into a live test against real policy variation. Each is a study in its own right; together they turn a single measurement of the price of children into a program for asking what that price is, for whom, and how policy might change it.

## References


**Adda, Jérôme, Christian Dustmann, and Katrien Stevens**, "The Career Costs of Children," *Journal of Political Economy*, 2017, *125* (2), 293–337.

**Adrjan, Pawel, Mária Balgova, Simon Jäger, Jonas Jessen, and Jason Sockin**, "Job Ads as Signals: Evidence from a Priced Amenity and Worker Beliefs," NBER Working Paper 35457, National Bureau of Economic Research 2026.

**Almås, Ingvild, Orazio Attanasio, and Pamela Jervis**, "Presidential Address: Economics and Measurement: New Measures to Model Decision Making," *Econometrica*, 2024, *92* (4), 947–978.

**Andersson, Gunnar, Jan M. Hoem, and Ann-Zofie Duvander**, "Social Differentials in Speed-Premium Effects in Childbearing in Sweden," *Demographic Research*, 2006, *14* (4), 51–70.

**Backhaus, Teresa, Clara Schäper, and Annekatrin Schrenker**, "Causal Misperceptions of the Part-Time Pay Gap," *Labour Economics*, 2023, *83*, 102396.

**Bar, Michael, Moshe Hazan, Oksana Leukhina, David Weiss, and Hosny Zoabi**, "Why Did Rich Families Increase Their Fertility? Inequality and Marketization of Child Care," *Journal of Economic Growth*, 2018, *23* (4), 427–463.

**Becker, Gary S.**, "An Economic Analysis of Fertility," in "Demographic and Economic Change in Developed Countries," Columbia University Press, 1960, pp. 209–240.

_ , "A Theory of the Allocation of Time," *Economic Journal*, 1965, *75* (299), 493–517.

_ **and H. Gregg Lewis**, "On the Interaction between the Quantity and Quality of Children," *Journal of Political Economy*, 1973, *81* (2), S279–S288.

**Behrman, Julia, Emily A. Marshall, and Florian Keusch**, "An Experimental Approach to Assessing Young Women's Childbearing Preferences: A Research Note on the United States," *Demography*, 2025, *62* (3), 773–785.

**Bergsvik, Janna, Agnes Fauske, and Rannveig Kaldager Hart**, "Can Policies Stall the Fertility Fall? A Systematic Review of the (Quasi-)Experimental Literature," *Population and Development Review*, 2021, *47* (4), 913–964.

**Berman, Eli, Laurence R. Iannaccone, and Giuseppe Ragusa**, "From Empty Pews to Empty Cradles: Fertility Decline among European Catholics," *Journal of Demographic Economics*, 2018, *84* (2), 149–187.

**Black, Dan A., Natalia Kolesnikova, Seth G. Sanders, and Lowell J. Taylor**, "Are Children "Normal"?," *Review of Economics and Statistics*, 2013, *95* (1), 21–33.

**Bokun, Anna**, "Cash Transfers and Fertility: Evidence from Poland's Family 500+ Policy," *Demographic Research*, 2024, *51* (28), 855–910.

**Bradbury, Bruce**, "Time and the Cost of Children," *Review of Income and Wealth*, 2008, *54* (3), 305–323.

**Cesarini, David, Erik Lindqvist, Robert Östling, and Anastasia Terskaya**, "Fortunate Families? The Effects of Wealth on Marriage and Fertility," Technical Report, NBER Working Paper No. 31039 2023.

**Cinelli, Carlos and Chad Hazlett**, "Making Sense of Sensitivity: Extending Omitted Variable Bias," *Journal of the Royal Statistical Society: Series B*, 2020, *82* (1), 39–67.

**Clarke, Damian, Joseph P. Romano, and Michael Wolf**, "The Romano–Wolf Multiple-Hypothesis Correction in Stata," *The Stata Journal*, 2020, *20* (4), 812–843.

**Cogan, John F.**, "Fixed Costs and Labor Supply," *Econometrica*, 1981, *49* (4), 945–963.

**Cortés, Patricia and Jessica Pan**, "Children and the Remaining Gender Gaps in the Labor Market," *Journal of Economic Literature*, 2023, *61* (4), 1359–1409.

**Craig, Lyn and Killian Mullan**, "Parenthood, Gender and Work-Family Time in the United States, Australia, Italy, France, and Denmark," *Journal of Marriage and Family*, 2010, *72* (5), 1344–1361.

_ **and Michael Bittman**, "The Incremental Time Costs of Children: An Analysis of Children's Impact on Adult Time Use in Australia," *Feminist Economics*, 2008, *14* (2), 59–88.

**Dafoe, Allan, Baobao Zhang, and Devin Caughey**, "Information Equivalence in Survey Experiments," *Political Analysis*, 2018, *26* (4), 399–416.

**Dagsvik, John K. and Anders Karlström**, "Compensating Variation and Hicksian Choice Probabilities in Random Utility Models that are Nonlinear in Income," *Review of Economic Studies*, 2005, *72* (1), 57–76.

**Dahl, Gordon B., Katrine V. Løken, Magne Mogstad, and Kari Vea Salvanes**, "What Is the Case for Paid Maternity Leave?," *The Review of Economics and Statistics*, 2016, *98* (4), 655–670.

**Delventhal, Matthew J., Jesús Fernández-Villaverde, and Nezih Guner**, "Demographic Transitions Across Time and Space," NBER Working Paper 29480, National Bureau of Economic Research 2021.

**Doepke, Matthias and Fabian Kindermann**, "Bargaining over Babies: Theory, Evidence, and Policy Implications," *American Economic Review*, 2019, *109* (9), 3264–3306.

_ **, Anne Hannusch, Fabian Kindermann, and Michèle Tertilt**, "The Economics of Fertility: A New Era," in "Handbook of the Economics of the Family," Vol. 1 2023, pp. 151–254.

**Fernández-Villaverde, Jesús and Patrick Norrick**, "Terra Incognita: The Economics of a Shrinking World," Technical Report, University of Pennsylvania and Northwestern University 2026.

**Flood, Sarah M., Liana C. Sayer, and Daniel Backman**, "American Time Use Survey Data Extract Builder: Version 3.3 [dataset]," College Park, MD: University of Maryland, and Minneapolis, MN: IPUMS 2025. `https://doi.org/10.18128/D060.V3.3`.

**Galor, Oded and David N. Weil**, "The Gender Gap, Fertility, and Growth," *American Economic Review*, 1996, *86* (3), 374–387.

**Geruso, Michael and Dean Spears**, "Childless or Child-fewer? Childlessness and Parity Progression Where Fertility Is Below Replacement," NBER Working Paper 33913, National Bureau of Economic Research 2025.

**Goldin, Claudia**, "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 2014, *104* (4), 1091–1119.

_ , *Career and Family: Women's Century-Long Journey toward Equity*, Princeton University Press, 2021.

**Goldstein, Joshua, Christos Koulovatianos, Jian Li, and Carsten Schroeder**, "Combining a Survey Experiment with Lifecycle Models to Evaluate Pronatalist Policies," Technical Report, DIW Berlin Discussion Paper No. 2171 2026.

**González, Libertad**, "The Effect of a Universal Child Benefit on Conceptions, Abortions, and Early Maternal Labor Supply," *American Economic Journal: Economic Policy*, 2013, *5* (3), 160–188.

_ **and Sofia Karina Trommlerová**, "Cash Transfers and Fertility: How the Introduction and Cancellation of a Child Benefit Affected Births and Abortions," *Journal of Human Resources*, 2023, *58* (3), 783–818.

**Gromadzki, Jan**, "Universal Child Benefit and Child Poverty: The Role of Fertility Adjustments," Technical Report, IZA Discussion Paper No. 17456 2024.

**Hagewen, Kellie J. and S. Philip Morgan**, "Intended and Ideal Family Size in the United States, 1970–2002," *Population and Development Review*, 2005, *31* (3), 507–527.

**Hamilton, Brady E., Joyce A. Martin, and Michelle J. K. Osterman**, "Births: Provisional Data for 2024," Vital Statistics Rapid Release 38, National Center for Health Statistics 2025.

**Hayford, Sarah R. and S. Philip Morgan**, "Religiosity and Fertility in the United States: The Role of Fertility Intentions," *Social Forces*, 2008, *86* (3), 1163–1188.

**Hotz, V. Joseph and Robert A. Miller**, "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply," *Econometrica*, 1988, *56* (1), 91–118.

**Inglehart, Ronald F.**, *Religion's Sudden Decline: What's Causing It, and What Comes Next?*, Oxford University Press, 2021.

**Kearney, Melissa S. and Phillip B. Levine**, "Why Is Fertility So Low in High-Income Countries?," NBER Working Paper 33989, National Bureau of Economic Research 2025.

**Kleven, Henrik, Camille Landais, and Gabriel Leite-Mariante**, "The Child Penalty Atlas," *Review of Economic Studies*, 2025, *92* (5), 3174–3207.

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

_ **,** _ **, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller**, "Child Penalties across Countries: Evidence and Explanations," *AEA Papers and Proceedings*, 2019, *109*, 122–126.

**Kuziemko, Ilyana, Jessica Pan, Jenny Shen, and Ebonya Washington**, "The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?," NBER Working Paper 24740, National Bureau of Economic Research 2018.

**Maestas, Nicole, Kathleen J. Mullen, David Powell, Till von Wachter, and Jeffrey B. Wenger**, "The Value of Working Conditions in the United States and the Implications for the Structure of Wages," *American Economic Review*, 2023, *113* (7), 2007–2047.

**Markussen, Simen and Marte Strøm**, "Children and Labor Market Outcomes: Separating the Effects of the First Three Children," *Journal of Population Economics*, 2022, *35* (1), 135–167.

**Mas, Alexandre and Amanda Pallais**, "Valuing Alternative Work Arrangements," *American Economic Review*, 2017, *107* (12), 3722–3759.

**Mincer, Jacob**, "Market Prices, Opportunity Costs, and Income Effects," in Carl Christ, ed., *Measurement in Economics: Studies in Mathematical Economics in Honor of Yehuda Grunfeld*, Stanford, CA: Stanford University Press, 1963.

**Musick, Kelly, Pilar Gonalons-Pons, and Christine R. Schwartz**, "Change and Variation in U.S. Couples' Earnings Equality Following Parenthood," *Population and Development Review*, 2022, *48* (2), 413–443.

**Oster, Emily**, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

**Pew Research Center**, "Decline of Christianity in the U.S. Has Slowed, May Have Leveled Off: 2023–24 Religious Landscape Study," Technical Report, Pew Research Center 2025.

**Romano, Joseph P. and Michael Wolf**, "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing," *Journal of the American Statistical Association*, 2005, *100* (469), 94–108.

**Schoen, Robert, Nan Marie Astone, Young J. Kim, Constance A. Nathanson, and Jason M. Fields**, "Do Fertility Intentions Affect Fertility Behavior?," *Journal of Marriage and the Family*, 1999, *61* (3), 790–799.

**Schrenker, Annekatrin**, "Do Women Expect Wage Cuts for Part-Time Work?," *Labour Economics*, 2023, *80*, 102291.

**Sheppard, Paula**, "Using Discrete Choice Modeling to Understand the Drivers of Reproductive Delay in the United Kingdom," *International Journal of Population Studies*, 2025, *11* (3), 125–137.

**Small, Kenneth A. and Harvey S. Rosen**, "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 1981, *49* (1), 105–130.

**Sobotka, Tomáš and Éva Beaujouan**, "Two Is Best? The Persistence of a Two-Child Family Ideal in Europe," *Population and Development Review*, 2014, *40* (3), 391–419.

**Stantcheva, Stefanie**, "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible," *Annual Review of Economics*, 2023, *15*, 205–234.

**Swait, Joffre and Jordan Louviere**, "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models," *Journal of Marketing Research*, 1993, *30* (3), 305–314.

**United Nations, Department of Economic and Social Affairs, Population Division**, "World Fertility 2024," Technical Report, United Nations 2025.

**U.S. Census Bureau**, "Real Median Household Income in the United States (MEHOINUSA672N)," Retrieved from FRED, Federal Reserve Bank of St. Louis 2025.

**Willis, Robert J.**, "A New Approach to the Economic Theory of Fertility Behavior," *Journal of Political Economy*, 1973, *81* (2), S14–S64.

**Wiswall, Matthew and Basit Zafar**, "Preference for the Workplace, Investment in Human Capital, and Gender," *Quarterly Journal of Economics*, 2018, *133* (1), 457–507.

_ **and** _ , "Human Capital Investments and Expectations about Career and Family," *Journal of Political Economy*, 2021, *129* (5), 1361–1424.

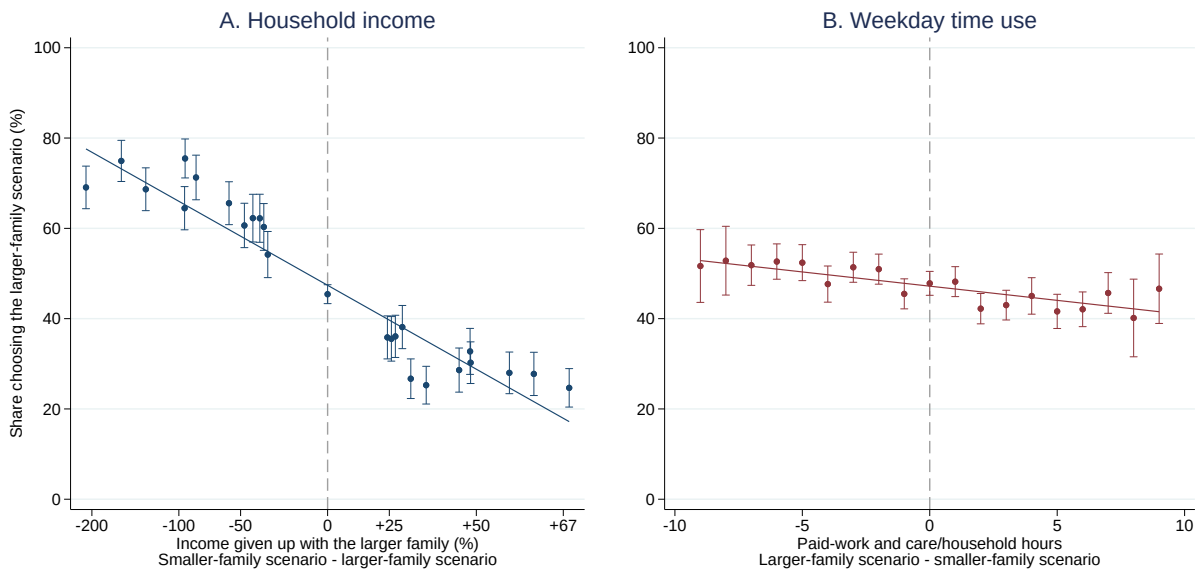



Figure 1: Demand for additional children, by money and time price

*Notes:* Points are the observed share choosing the scenario with more additional children at each fielded price, with 95 percent intervals from standard errors clustered on the respondent.
The line is the fitted value from the pooled reduced form reported in Table 3 column (1),

$$y_{it} = \delta_i + \theta P_{it} + \phi T_{it} + \lambda L_{it} + u_{it},$$

where $P$ is the money price, the difference in log household income between the smaller-family and larger-family scenarios, $T$ is the time price, the difference in non-leisure hours, and $L$ indicates that the larger-family scenario appeared on the left. Each panel varies one price and holds the other at its mean.
The respondent fixed effects absorb the intercept, so the height of the line is not estimated; it is drawn through the sample mean. Points and line are therefore not conditioned on the same things, since the raw shares average over the other price and over respondents.

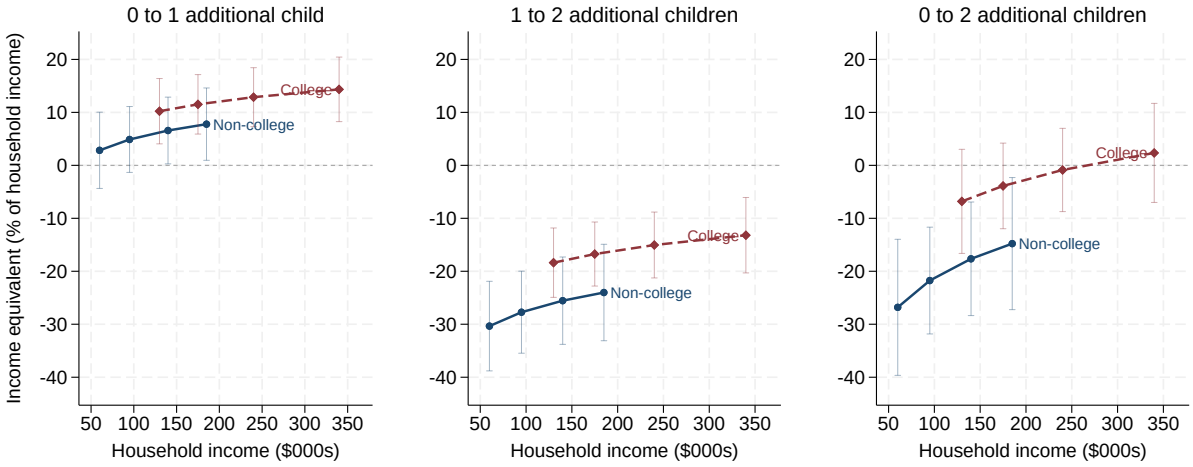
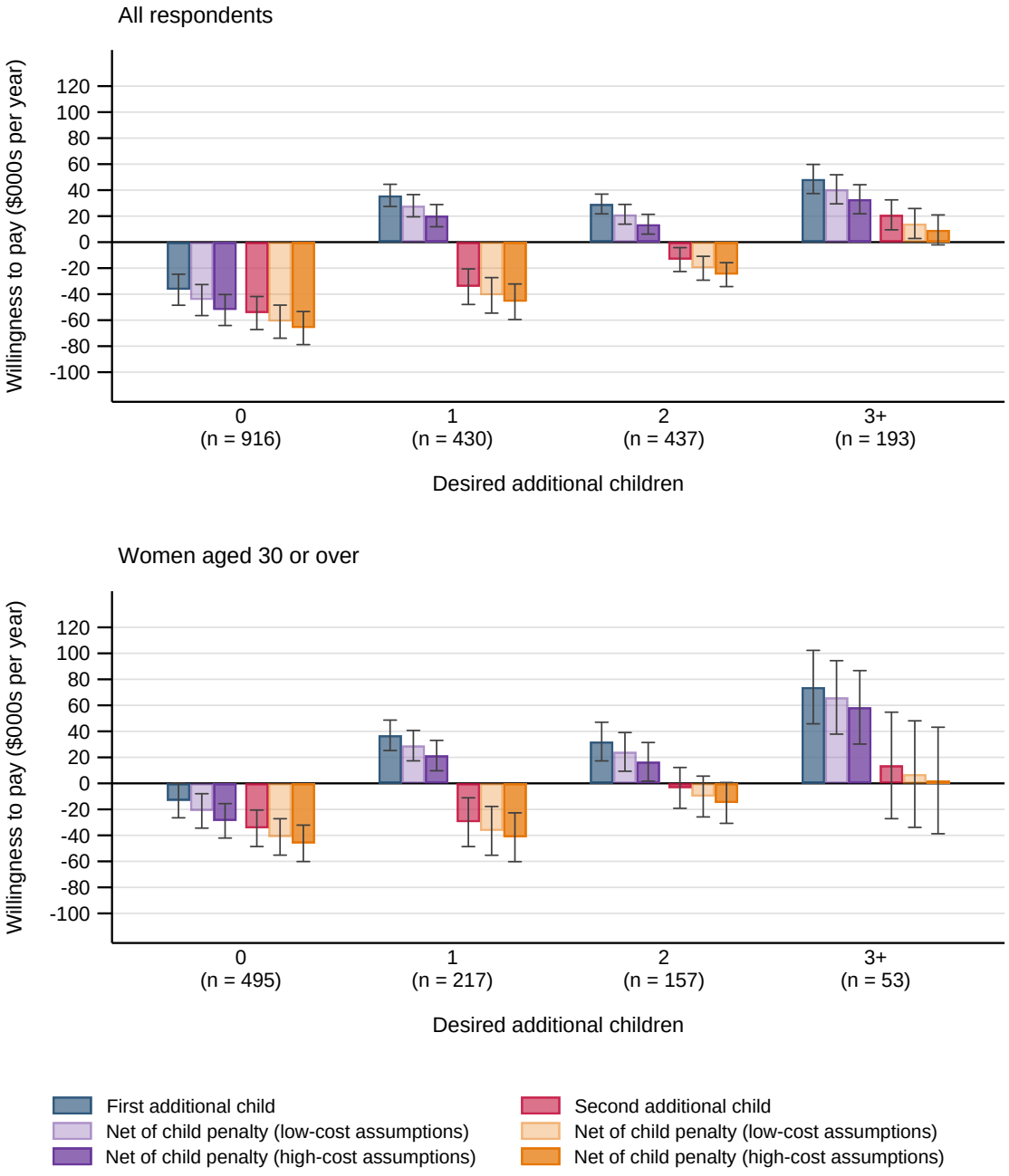



Figure 2: Willingness to pay as a share of household income, by child comparison

*Notes:* Willingness to pay as a percentage of household income at each of the four income levels shown to each education group, from the conditional logits of Table 5, estimated separately for each education group. Solid circles denote non-college respondents and dashed diamonds denote college respondents. Bars are 95 percent confidence intervals. A positive value is the share of income a respondent would give up for the larger family, and a negative value is the compensation that family must provide. A profile that rises with income indicates that the valuation grows more than proportionally with income. These estimates illustrate the registered reduced-form H2 test of Section 4.3 but do not replace it.

Figure 3: Willingness to pay for a first and second additional child, by desired additional fertility




*Notes:* Willingness to pay at a household income of $125,000 a year, from the conditional logit estimated separately by desired additional fertility. The upper panel uses all respondents and the lower panel women aged 30 or over. Numbers in parentheses are the respondents in each group. Blue and pink bars show willingness to pay for the first and second additional child. Purple and orange bars subtract the child penalty, the annual value of the fall in the mother's earnings and the rise in her hours of paid and unpaid work after a birth, under low (light) and high (dark) estimates: $7,900 to $15,600 for a first additional child and $6,700 to $11,600 for a second. Goods costs are not subtracted, because respondents pay for them out of the household income stated in each scenario. Because the penalty is a fixed amount, the intervals on the net bars are those of the unadjusted bars shifted by it, and do not include uncertainty in the penalty estimates. Capped lines show 95 percent confidence intervals clustered on the respondent.

Table 1: Household-income levels shown in the experiment

| Education group | 25th percentile | 50th percentile | 75th percentile | 90th percentile |
|---|---|---|---|---|
| Without a college degree | $60,000 | $95,000 | $140,000 | $185,000 |
| College graduate | $130,000 | $175,000 | $240,000 | $340,000 |

Table 2: The realized sample against population benchmarks

| *Panel A. Margins used to steer recruitment* | | | | |
|---|---|---|---|---|
| | Respondents | Sample | Population | Difference |
| College | 936 | 46.9 | 45.2 | +1.7 |
| Aged under 30 | 1,061 | 53.2 | 52.1 | +1.1 |
| Non-Hispanic white alone | 1,002 | 50.8 | 50.8 | −0.1 |
| Metropolitan | 1,631 | 83.1 | 88.6 | −5.5 |
| Parent | 923 | 46.5 | 46.4 | +0.1 |
| *Panel B. Post-stratification cells* | | | | |
| | Respondents | Sample | Population | Weight |
| Non-college, under 30 | 595 | 29.8 | 29.2 | 0.980 |
| Non-college, 30 and over | 464 | 23.3 | 25.6 | 1.102 |
| College, under 30 | 466 | 23.4 | 22.8 | 0.977 |
| College, 30 and over | 470 | 23.6 | 22.3 | 0.948 |
| *Panel C. Design strata and the cost of weighting* | | | | |
| Non-college | 1,059 | 53.1 | | |
| College | 936 | 46.9 | | |
| Design effect | | | | 1.003 |
| Effective sample size | | | | 1,988 |

*Notes:* Primary sample, 1,995 respondents, unweighted. Shares are in percent. The ACS benchmarks use women aged 23–35 in the 2024 one-year file; the parenthood benchmark uses the June 2024 CPS Fertility Supplement. Education and age are available for all respondents. The remaining Panel A shares use the answered denominator: 1,963 for urban–rural status, 1,974 for non-Hispanic White-alone status and 1,985 for parenthood. Panel B weights equal the population cell share divided by the realized sample share. Panel C reports the two education strata and the variance cost of weighting. Appendix B.5 describes the benchmark construction and reconciles differences between the survey and population definitions.

Table 3: Demand for a larger family

| | (1) Baseline (H1, H3) | (2) Income interaction (H2) | (3) Time split |
|---|---|---|---|
| *Panel A. Estimated coefficients* | | | |
| Money price $P_{it}$ | −0.268 | −0.239 | −0.268 |
| | (0.009) | (0.012) | (0.009) |
| | [$p < 0.001$] | | |
| Time price $T_{it}$ | −0.0063 | −0.0063 | |
| | (0.0011) | (0.0011) | |
| | [$p < 0.001$] | | |
| Paid-work hours $T_{it}^W$ | | | −0.0063 |
| | | | (0.0012) |
| Care and household hours $T_{it}^H$ | | | −0.0065 |
| | | | (0.0016) |
| Additional children $\Delta k_{it}$ | | −0.003 | |
| | | (0.008) | |
| Additional children × log income $C_{it}$ | | 0.028 | |
| | | (0.008) | |
| | | [0.012, 0.045] | |
| *Panel B. Value of an hour, percent of household income* | | | |
| All non-leisure hours, $\phi/\theta$ | 2.35 | | |
| | (0.43) | | |
| Paid-work hours | | | 2.34 |
| | | | (0.44) |
| Care and household hours | | | 2.41 |
| | | | (0.60) |
| Tasks | 13,442 | 13,442 | 13,442 |
| Respondents | 1,995 | 1,995 | 1,995 |

*Notes:* Let $y_{it} = 1$ if respondent $i$ chooses the scenario with more additional children in task $t$. Let $P_{it}$ be the difference in log household income between the fewer- and the more-children scenario, so $P_{it} > 0$ when the larger family is the more expensive one, and let $T_{it}$ be the difference in total non-leisure hours. Column (1) estimates the linear probability model

$$y_{it} = \delta_i + \theta P_{it} + \phi T_{it} + \lambda L_{it} + u_{it},$$

where $\delta_i$ is a respondent fixed effect and $L_{it}$ indicates that the larger-family scenario appeared on the left. Column (3) replaces $T_{it}$ with its paid-work and care components $T_{it}^W$ and $T_{it}^H$.

Column (2) adds two terms. $\Delta k_{it}$ is the difference in additional children between the two scenarios, one or two, and $C_{it}$ is the same difference interacted with log income centered on its sample mean,

$$C_{it} = k_{mt}\left(\ln Y_{mt} - \overline{\ln Y}\right) - k_{ft}\left(\ln Y_{ft} - \overline{\ln Y}\right),$$

writing $m$ and $f$ for the more- and fewer-children scenarios. The coefficient on $C_{it}$ is $\mu$, the registered H2 estimand: it is positive when a larger family becomes more attractive as income rises. $\Delta k_{it}$ is included because an interaction requires its own main effect; its coefficient is the propensity to choose the larger family per additional child on offer, at equal money and time prices and at mean income, and is not itself of interest.

All variables are differences across the two scenarios in a task, so every coefficient is identified within respondent. The estimation sample is the tasks in which the two scenarios differ in additional children. Standard errors in parentheses, clustered on the respondent. Square brackets carry the registered inference and appear only on the three pre-registered parameters: a one-sided test for each price, as the sign of both was predicted in advance, and a 95 percent interval for $\mu$.

Panel B reports the income a respondent would need in exchange for one weekday hour moved out of leisure and sleep, as a percentage of household income. It is $\phi/\theta$, and the paid and care equivalents from column (3), with delta-method standard errors.

Table 4: Non-parametric estimates of the time price

| | (1) Linear time price | (2) Hour levels |
|---|---|---|
| *Panel A. Price coefficients* | | |
| Money price $P_{it}$ | −0.268 | −0.267 |
| | (0.009) | (0.009) |
| Time price $T_{it}$ | −0.0063 | |
| | (0.0011) | |
| Paid work, 4 hours | | 0.017 |
| | | (0.009) |
| Paid work, 9 hours | | −0.043 |
| | | (0.011) |
| Care and household tasks, 3 hours | | −0.002 |
| | | (0.008) |
| Care and household tasks, 5 hours | | −0.019 |
| | | (0.008) |
| Care and household tasks, 8 hours | | −0.029 |
| | | (0.010) |
| *Panel B. Value of each hour level, percent of household income* | | |
| Paid work, 4 hours | | 6.53 |
| | | (3.48) |
| Paid work, 9 hours | | −14.74 |
| | | (3.48) |
| Care and household tasks, 3 hours | | −0.73 |
| | | (3.14) |
| Care and household tasks, 5 hours | | −7.01 |
| | | (2.84) |
| Care and household tasks, 8 hours | | −10.13 |
| | | (3.37) |
| Step from 4 to 9 hours | | −19.97 |
| | | (2.32) |
| $p$-value for H0: paid-work effects proportional to hours | | $< 0.001$ |
| $p$-value for H0: care effects proportional to hours | | 0.59 |
| $p$-value for H0: paid and care effects equal at equal hours | | $< 0.001$ |
| Tasks | 13,442 | 13,442 |
| Respondents | 1,995 | 1,995 |

*Notes:* Column (1) is the registered specification of Table 3, in which the time price is the single linear term $T_{it}$, the difference in total non-leisure hours. Column (2) imposes no shape on hours. Writing $m$ and $f$ for the more- and fewer-children scenarios, it estimates

$$y_{it} = \delta_i + \theta P_{it} + \sum_{h \in \{4,9\}} \phi_h^W \left(\mathbf{1}\{W_{mt} = h\} - \mathbf{1}\{W_{ft} = h\}\right) + \sum_{h \in \{3,5,8\}} \phi_h^H \left(\mathbf{1}\{H_{mt} = h\} - \mathbf{1}\{H_{ft} = h\}\right) + \lambda L_{it} + u_{it},$$

so each fielded hour level has its own coefficient. The omitted levels are zero hours of paid work and two hours of care, which are therefore normalized to zero rather than estimated. Both columns include respondent fixed effects and are estimated on tasks in which the two scenarios differ in additional children. Standard errors in parentheses, clustered on the respondent. Panel B converts each coefficient into the proportional income change that offsets it, $\exp(-b/\theta) - 1$, with delta-method standard errors. The step from four to nine hours of paid work is estimated as a linear combination of the two level coefficients rather than by differencing them: the point estimate is the difference, but its standard error requires their covariance, which corresponds to a correlation of 0.72, and ignoring it would overstate the standard error by close to a factor of two.

Table 5: Willingness to pay for additional children

| Additional children | Non-college at median income $95,000 | College at median income $175,000 |
|---|---|---|
| One versus none | $4,643 | $20,162 |
| | [−1,263, 10,550] | [10,333, 29,990] |
| Two versus one | −$26,337 | −$29,301 |
| | [−33,696, −18,977] | [−39,880, −18,722] |
| Two versus none | −$20,662 | −$6,797 |
| | [−30,236, −11,089] | [−20,939, 7,344] |

*Notes:* Conditional logit in Equation (7), estimated separately for each education stratum on all primary tasks. Each entry is the change in annual household income that leaves a respondent indifferent between the two family sizes, evaluated at the stratum median income. A positive figure is income the respondent would give up for the larger-family scenario; a negative figure is the compensation that scenario must provide to leave her as well off. Brackets contain 95 percent confidence intervals calculated by the delta method, with standard errors clustered on the respondent. Appendix Table A5 reports the same quantity at every fielded income level, and Appendix Table A6 reports it in weekday hours.

Table 6: Willingness to pay for additional children, by current parity

| | Respondents | First additional child | Second additional child |
|---|---|---|---|
| *Panel A. Non-college women, at $95,000* | | | |
| Childless | 553 | −$16,153 | −$47,719 |
| | | [−26,468, −5,838] | [−60,509, −34,928] |
| Mothers | 500 | $22,406 | −$6,901 |
| | | [15,738, 29,073] | [−15,034, 1,232] |
| *Panel B. College women, at $175,000* | | | |
| Childless | 509 | −$11,826 | −$44,378 |
| | | [−27,307, 3,656] | [−58,355, −30,401] |
| Mothers | 423 | $54,270 | −$8,829 |
| | | [41,790, 66,751] | [−25,534, 7,877] |

*Notes:* Exploratory estimates from the primary sample. The structural conditional logit in Equation (7) is estimated separately within each of the four subgroups formed by education stratum and motherhood status, rather than as a pooled interaction model. All coefficients on additional children, log household income, the child-by-income interaction, paid-work hours, care and household hours, and the left-position indicator are therefore allowed to differ across subgroups. Childless women report no current children; mothers report at least one. Respondents with missing current parity are excluded, and no additional comprehension restriction is imposed. The table reports exact compensating income changes at the median displayed income in each education stratum, $95,000 for non-college women and $175,000 for college women. A positive estimate is annual household income the respondent would give up for the larger-family scenario; a negative estimate is the annual compensation that scenario must provide to leave her as well off. Brackets contain 95 percent confidence intervals calculated by the delta method, with standard errors clustered on the respondent. First and second refer to the first and second additional child, not necessarily the first and second birth.

Table 7: Heterogeneity in baseline demand and price responses

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *Panel A. Religiosity* | | | | |
| Baseline demand | 0.074 | | | 0.063 |
| | (0.006) | | | (0.006) |
| | [<0.001] | | | [<0.001] |
| Interaction with the money price | 0.012 | | | 0.016 |
| | (0.008) | | | (0.008) |
| | [0.095] | | | [0.097] |
| Interaction with the time price | 0.0042 | | | 0.0045 |
| | (0.0010) | | | (0.0010) |
| | [<0.001] | | | [<0.001] |
| *Panel B. Desired additional fertility* | | | | |
| Baseline demand | | 0.081 | | 0.079 |
| | | (0.006) | | (0.006) |
| | | [<0.001] | | [<0.001] |
| Interaction with the money price | | −0.022 | | −0.024 |
| | | (0.008) | | (0.008) |
| | | [0.001] | | [<0.001] |
| Interaction with the time price | | −0.0009 | | −0.0017 |
| | | (0.0009) | | (0.0009) |
| | | [0.255] | | [0.121] |
| *Panel C. Age 30 or older* | | | | |
| Baseline demand | | | 0.036 | 0.070 |
| | | | (0.014) | (0.013) |
| | | | [0.027] | [<0.001] |
| Interaction with the money price | | | 0.014 | 0.002 |
| | | | (0.019) | (0.019) |
| | | | [0.477] | [0.902] |
| Interaction with the time price | | | −0.0023 | −0.0030 |
| | | | (0.0023) | (0.0024) |
| | | | [0.477] | [0.301] |
| Observations | 13,098 | 13,098 | 13,098 | 13,098 |
| Respondents | 1,941 | 1,941 | 1,941 | 1,941 |

*Notes:* Each column combines two linear probability models estimated on the same sample. The baseline-demand rows report $\rho$ from

$$y_{it} = \delta_{c(i)} + \theta P_{it} + \phi T_{it} + M_i'(\pi_P P_{it} + \pi_T T_{it}) + M_i'\rho + \pi_C C_i P_{it} + \lambda L_{it} + u_{it},$$

and the interaction rows report $\pi_P$ and $\pi_T$ from

$$y_{it} = \delta_i + \theta P_{it} + \phi T_{it} + M_i'(\pi_P P_{it} + \pi_T T_{it}) + \pi_C C_i P_{it} + \lambda L_{it} + u_{it},$$

where $y_{it}$ equals one if respondent $i$ chooses the scenario with more children in task $t$, $P_{it}$ is the money price (log income given up), $T_{it}$ is the time price (weekday hours of paid and care work added), $C_i$ indicates the college stratum, $L_{it}$ indicates that the scenario with more children appears on the left, $\delta_{c(i)}$ is an education-stratum fixed effect, and $\delta_i$ is a respondent fixed effect, which absorbs the level term $M_i'\rho$. The vector $M_i$ contains one characteristic in columns 1–3 and all three in column 4. Each baseline-demand coefficient is the difference in the probability of choosing the scenario with more children when the two scenarios have the same income and hours. Standard errors clustered by respondent in parentheses. Romano–Wolf stepdown $p$-values in brackets, from 1,000 respondent-level bootstrap replications; each column is one family of tests, with three coefficients in columns 1–3 and nine in column 4. Religiosity is the importance of religion, from 1 (not at all important) to 4 (very important), and desired additional fertility is the ideal minus the current number of children; coefficients are per point, per child, and for age 30 or older versus younger. The moderators are not randomized, so the differences are descriptive; desired fertility is a stated intention and partly overlaps the outcome.

Table 8: Willingness to pay for additional children, by respondent characteristic

| | Respondents | First additional child | Second additional child |
|---|---|---|---|
| *Panel A. Non-college women, at $95,000* | | | |
| Religiosity | | | |
| Not important (1–2) | 453 | −$16,485 | −$40,319 |
| | | [−26,467, −6,504] | [−52,264, −28,374] |
| Important (3–4) | 587 | $20,015 | −$16,012 |
| | | [12,973, 27,057] | [−25,472, −6,553] |
| Desired additional fertility | | | |
| None wanted | 524 | −$27,604 | −$47,322 |
| | | [−39,454, −15,754] | [−61,684, −32,960] |
| One or more wanted | 526 | $25,991 | −$10,507 |
| | | [19,904, 32,079] | [−18,485, −2,530] |
| Age | | | |
| Under 30 | 595 | $2,231 | −$25,626 |
| | | [−5,563, 10,025] | [−34,955, −16,296] |
| 30 or over | 464 | $7,410 | −$27,025 |
| | | [−1,661, 16,481] | [−38,867, −15,183] |
| *Panel B. College women, at $175,000* | | | |
| Religiosity | | | |
| Not important (1–2) | 419 | −$14,797 | −$49,069 |
| | | [−30,318, 723] | [−63,609, −34,529] |
| Important (3–4) | 498 | $47,916 | −$10,067 |
| | | [35,287, 60,545] | [−25,394, 5,259] |
| Desired additional fertility | | | |
| None wanted | 392 | −$49,870 | −$65,062 |
| | | [−75,895, −23,844] | [−89,864, −40,261] |
| One or more wanted | 534 | $51,342 | −$14,983 |
| | | [42,410, 60,275] | [−26,353, −3,613] |
| Age | | | |
| Under 30 | 466 | $10,756 | −$36,997 |
| | | [−2,716, 24,228] | [−51,205, −22,790] |
| 30 or over | 470 | $30,705 | −$20,762 |
| | | [16,355, 45,054] | [−36,688, −4,837] |

*Notes:* Exploratory estimates from the primary sample, in the same units and at the same incomes as Table 6. The structural conditional logit in Equation (7) is fitted separately within each education-by-subgroup cell rather than as a pooled interaction model, so every coefficient, including the logit scale, is free to differ across the split. Willingness to pay is a ratio of two coefficients from the same fit, so the scale cancels and the rows are comparable.

Only the two adjacent steps are shown. Each subgroup fit has two free child parameters, $\alpha_1$ and $\alpha_2$, and all three child contrasts are functions of that pair, so a zero-to-two column would carry no information beyond the two columns here.

Religiosity is the analysis variable built in the data preparation, which runs from 1 not at all important to 4 very important and reverses the fielded item; the split is at 3, the most balanced binary the four-point scale allows. Desired additional fertility is the stated ideal minus current children, split at zero, so the small number of women reporting more children than they say they want sit in the wants-no-more row. Respondents missing a moderator are excluded from that pair of rows only, which is why the row counts do not sum to the stratum total. Brackets contain 95 percent confidence intervals by the delta method, with standard errors clustered on the respondent.

Table 9: Registered robustness checks

| Sample or specification | H1 Money price $\theta$ | H2 Income interaction $\mu$ | H3 Time price $\phi$ | Tasks | Respondents |
|---|---|---|---|---|---|
| Primary | −0.2682 | 0.0284 | −0.00629 | 13,442 | 1,995 |
| | (0.0092) | (0.0084) | (0.00115) | | |
| Post-stratification weighted | −0.2677 | 0.0282 | −0.00622 | 13,442 | 1,995 |
| | (0.0093) | (0.0085) | (0.00115) | | |
| Passed attention check | −0.2756 | 0.0293 | −0.00720 | 12,534 | 1,864 |
| | (0.0095) | (0.0086) | (0.00118) | | |
| Passed all comprehension checks | −0.3063 | 0.0349 | −0.00816 | 3,465 | 517 |
| | (0.0168) | (0.0154) | (0.00225) | | |
| Not straight-lined | −0.2711 | 0.0269 | −0.00624 | 13,281 | 1,971 |
| | (0.0093) | (0.0085) | (0.00116) | | |
| Slower two terciles by conjoint time | −0.2901 | 0.0396 | −0.00879 | 8,818 | 1,319 |
| | (0.0111) | (0.0106) | (0.00144) | | |
| Slower two terciles by core time | −0.2841 | 0.0394 | −0.00930 | 8,914 | 1,329 |
| | (0.0111) | (0.0104) | (0.00142) | | |
| Found scenarios realistic | −0.2701 | 0.0299 | −0.00500 | 9,069 | 1,336 |
| | (0.0114) | (0.0104) | (0.00142) | | |
| Excluding five anomalous responses | −0.2687 | 0.0284 | −0.00632 | 13,407 | 1,990 |
| | (0.0092) | (0.0085) | (0.00115) | | |

*Notes:* Each row applies the indicated weighting or sample restriction while otherwise leaving the relevant specification unchanged. H1 is the money-price coefficient $\theta$, H2 the child-by-log-income interaction $\mu$, and H3 the time-price coefficient $\phi$.
Two regressions lie behind each row because estimating $\mu$ requires terms that the price specification does not contain. Both regressions use the same respondents and tasks:

$$y_{it} = \delta_i + \theta P_{it} + \phi T_{it} + \lambda L_{it} + u_{it},$$

$$y_{it} = \delta_i + \theta P_{it} + \phi T_{it} + \kappa\, \Delta k_{it} + \mu\, C_{it} + \lambda L_{it} + u_{it},$$

where $\Delta k_{it}$ is the difference in additional children, $C_{it}$ is the child-by-log-income interaction centered on mean log income, and $L_{it}$ indicates that the larger-family scenario appears on the left. The table takes $\theta$ and $\phi$ from the first regression and $\mu$ from the second. Standard errors in parentheses are clustered by respondent.
Comprehension is measured using respondents' first attempts, before they received corrective feedback. The two timing restrictions retain the slower two-thirds of respondents according to time spent on the ten choice tasks alone and time spent on the instructions and choice tasks together, respectively. Appendix H defines the remaining restrictions and reports the accompanying analyses.

## Online Appendix for Money, Time, and the Demand for Children

### A Fertility intentions and completed fertility

To evaluate the gap between fertility intentions and completed fertility, we use the National Survey of Family Growth. We pair each of the four National Survey of Family Growth cycles fielded between 1982 and 2002 with a cycle fielded roughly twenty years later.[10] The results in Figure A1 show that intended family size exceeds completed fertility in every cohort; across the four, the gap averages 0.34 children.

### B Data appendix

#### *B.1 Panel, recruitment, and fieldwork*

We recruited 1,995 women living in the United States through Bilendi, an online survey panel, between 22 July and 17 August 2026. Bilendi's members volunteer to take part in research and receive points redeemable for cash or vouchers. Our survey was offered only to registered panel members, and Bilendi used its registration and entry checks to reduce the risk of duplicate participation.

Before fieldwork, we gave Bilendi target shares for education, age, area of residence, race and ethnicity, and parity. As responses accumulated, Bilendi monitored the composition of the sample and gave greater priority to groups that remained below their target shares.

Eligibility required respondents to identify as women, live in the United States, and have been born between 1990 and 2003. We used birth year rather than an exact date so that we could screen on age without collecting respondents' full dates of birth. These cohorts correspond to the target age range of 23–35 in 2026. Because exact dates of birth were not collected, respondents born in 1990 may have been aged 35 or 36 when they completed the survey, while those born in 2003 may have been aged 22 or 23. Both boundary cohorts are retained.

Respondents whose total survey time exceeded four minutes received points worth approximately \$1.90–\$2.15. We set this minimum before main fieldwork and registered it as the rule defining the primary sample. Median completion time for the 1,995 respondents was 446 seconds, or 7.4 minutes, making the payment equivalent to approximately \$15.30–\$17.40 per hour.

To finalize the instrument, we first conducted a separate pilot of 153 respondents. We used the pilot evidence to set the education-specific income grids, revise the survey instructions, and refine the comprehension checks.

#### *B.2 Primary and analysis-specific samples*

We intended to recruit 2,000 respondents who answered all ten choice questions and exceeded the registered four-minute threshold. The fieldwork counter reached 2,000, but it counted each

[10] Within each pair, both samples are restricted to the same approximate birth-year band, defined as interview year minus completed age. Estimates use each cycle's final survey weight. Standard errors use the survey-design variables or replicate weights released with each cycle. Each cohort-specific shortfall is statistically different from zero at the 5 percent level.

time the final page registered a completion rather than the number of distinct respondents. Six additional records were generated when four panel members refreshed the final page. One genuine completion generated no final-page record because the respondent did not click through from that page. Correcting for these cases left 1,995 distinct respondents in the primary sample. A further 299 respondents answered all ten choice questions in four minutes or less and were excluded under the registered rule. The primary sample therefore contains 19,950 choices between 39,900 scenario profiles.

The target of 2,000 respondents was fixed before fieldwork as the largest sample the panel could supply from the eligible population. Power calculations based on the pilot, reported in the pre-analysis plan, indicated that a sample of this size would detect money- and time-price responses well below those estimated in the pilot.

We selected the four-minute threshold after examining completion times in the pilot. It lies above the time needed simply to progress through the questionnaire and below the completion time of a typical respondent.

Most of the reduced-form analysis asks how the relative income and time use of the two scenarios affect demand for additional children. These estimates use the 13,442 tasks in which the scenarios offer different numbers of children; every respondent in the primary sample contributes at least one such task. The structural analysis instead uses all 19,950 tasks, since choices between scenarios offering the same number of children still reveal how respondents value income and time.

The analyses of stated fertility preferences also require answers to the questions on respondents' current and ideal numbers of children. Nineteen respondents omitted at least one of these questions, leaving 1,976 respondents and 13,319 unequal-child tasks for analyses that use the resulting measure of desired additional fertility. Further restrictions are specific to the comparison being made: for example, an analysis of whether respondents choose the family size they say they want uses only tasks that offer that family size.

Bilendi assigns each panel member a unique identifier, which allowed us to deal consistently with repeated survey attempts. One respondent completed the choice survey twice. We retained her first response because her later choices could have been affected by already having seen the instrument. Another respondent left the survey before reaching the choice questions and returned the following day to complete it. We retained the completed response because the earlier visit contained no choice data.

Two further cases required a decision. In neither does the survey record establish what happened, and both meet the registered timing rule, so both remain in the primary sample.

One respondent was screened out as male in July, re-entered a minute later giving their gender as female, and completed the survey. Gender is an eligibility criterion rather than a measure of response quality, and the two answers contradict each other, so eligibility cannot be established from the data. This is the only record for which that is true.

Four panel identifiers entered within two and a half hours of each other on a single day, share a browser fingerprint, and give the same answers to all nine questions the choice survey asks that the

gateway never saw: current children, ideal children, area, ethnicity, earnings, religion, religiosity, relationship and occupation. Agreement on nine independent answers is what one person completing the survey four times would produce, and not what four people sharing a device would produce. All four failed both the attention and the comprehension checks, and three spent between 42 and 45 seconds on all ten choice tasks. Panel identifier is the unit the provider supplies, and nothing in the instrument establishes that these are one person.

We retain all five. The registered plan defines the primary sample by the four-minute threshold alone, and every other response-quality measure enters as a robustness restriction rather than an exclusion; removing these records after observing them would substitute our judgment for that rule. Table 9 reports the confirmatory estimates with all five dropped.

### *B.3 Respondent characteristics and missingness*

The survey records the respondent characteristics used to describe the sample and examine how demand varies across women. Before the choice tasks, respondents report their birth year, gender, employment status, education, and whether they live in the United States. After the tasks, they are asked about their current and ideal numbers of children, relationship status, occupation, religion, the importance of religion in their lives, urban–rural status, race and ethnicity, and their own earnings during the previous year. The occupation, religion, and race and ethnicity questions each provide an "other" write-in option, and respondents may select more than one race or ethnicity. An optional text box at the end asks respondents to describe how they made their choices. We classify respondents as living with a partner if they are married, in a civil union, or living with a committed partner. Urban–rural status distinguishes cities and suburbs from rural areas and small towns, while the White-alone measure identifies respondents who selected White and no other race or ethnicity category. College status identifies respondents with a four-year degree or a postgraduate qualification.

The earnings question concerns the respondent's own earnings from work during the previous year, rather than the future combined household income shown in the choice tasks. The four earnings bands are \$0–\$9,999, \$10,000–\$34,999, \$35,000–\$59,999, and \$60,000 or more; respondents could also select "don't know" or "prefer not to answer."

Current and ideal numbers of children are top-coded at six or more. We construct desired additional fertility by subtracting a respondent's current number from her ideal number. Forty-seven respondents selected the top category on at least one of the two questions. For these respondents, the constructed difference may not equal the exact number of additional children they would ideally like. The relevant analyses are therefore repeated after excluding these respondents as a robustness check.

The five screening questions were required, so birth year, gender, employment status, education, and US residence are available for every respondent in the primary sample. This is because birth year, gender, and US residence determine eligibility, while education determines which income grid a respondent sees; employment status was collected in the same short screening block. Following

Stantcheva (2023), we placed the more personal questions after the choice tasks and allowed respondents to skip them. An unwillingness to answer a personal question therefore produces item non-response rather than attrition from the experimental sample.

Nineteen respondents omitted at least one of the current- and ideal-children questions, four of them both, so 1,976 answered both and can enter analyses using desired additional fertility. Ten respondents left race and ethnicity blank and eleven selected "prefer not to answer." Excluding both groups leaves 1,974 respondents for the constructed White-alone measure. Analyses using the remaining characteristics use the respondents who answered each item. The earnings item was answered by 1,992 respondents, of whom 43 selected "don't know" and 59 selected "prefer not to answer." Where earnings enter the analysis, these responses are retained as separate categories rather than treated as earnings bands.

### *B.4 Comprehension and response diagnostics*

The survey also includes three comprehension checks assessing respondents' understanding of the children, income, and time-use attributes. Respondents who answered incorrectly received feedback before continuing, so their first answers provide a measure of initial comprehension while the correction reinforces the survey instructions. On the first attempt, 83.8 percent answered the children question correctly, 75.5 percent answered the income question correctly, and 39.9 percent answered the time-use question correctly; 25.9 percent answered all three correctly.

We use an attention check and respondents' choice patterns as two further measures of engagement. The attention check was answered correctly by 93.4 percent of the primary sample, and 24 respondents, or 1.2 percent, chose the scenario on the same side of the screen in all ten tasks. The randomized design also creates a dominance check when two scenarios offer the same number of children and the same time allocation but different incomes. Because no other displayed attribute differs, the scenario with the higher income dominates the alternative under the maintained assumption that respondents prefer more income. Of the 484 respondents who faced at least one such task, 44, or 9.1 percent, chose the lower-income scenario at least once.

We report the distribution of total survey completion time and two measures that isolate time spent on the experimental material. Total completion time also includes the screening, instructions, post-task questions, and any period in which a respondent left the survey open before returning to it. The two narrower measures record time on the ten choice tasks alone and on the instructions and choice tasks together (Table A1).

Each of these two measures defines a robustness sample comprising the slower two-thirds of its timing distribution.

Perceived realism provides a complementary assessment of how respondents viewed the scenarios. Among the 1,994 respondents who answered the realism question, 1,336 reported that at least a majority of the scenarios felt realistic. The robustness section compares the primary estimates with estimates from this group and from the timing, comprehension, attention, and straight-lining samples described above.

*B.5 Sample composition and weighting*

To assess the composition of the realized sample, we compare the non-probability panel with population benchmarks on the five characteristics used to guide recruitment. Four margins are close to their population shares: college education differs by 1.7 percentage points, age by 1.1 points, and both White-alone status and parenthood by 0.1 point or less. The sample contains a smaller urban share than the benchmark, by 5.5 percentage points (Table 2, Panel A).

The two education strata contain 1,059 non-college respondents and 936 college respondents.

The education, age, race and ethnicity, and urban–rural benchmarks come from the 2024 one-year American Community Survey for women aged 23–35. The ACS metropolitan-status variable is indeterminate for 12.6 percent of age-eligible records because some Public Use Microdata Areas (PUMAs) include both metropolitan and nonmetropolitan areas. Simply excluding these records would produce a metropolitan share of 93.0 percent among the remaining observations. We instead use the share of each PUMA's 2020 Census population living in a 2023 metropolitan area, obtained from the Missouri Census Data Center's Geocorr correspondence, to allocate these observations fractionally. This gives the metropolitan benchmark of 88.6 percent. We use the June 2024 Current Population Survey Fertility Supplement for parenthood because it records whether a woman has ever given birth and therefore includes biological children who no longer live with her, whereas the ACS records only children living in her household. The sample shares for the optional post-task questions use the answered denominators reported in the notes to Table 2.

The urban–rural comparison is approximate because the ACS metropolitan measure is broader than the city-or-suburb classification used in the survey. The survey follows the 2024 federal standard and records Middle Eastern or North African separately, whereas the 2024 ACS includes these origins within White. This distinction is quantitatively small: treating respondents who selected only Middle Eastern or North African, or that category together with White, as White raises the realized White share from 50.8 to 51.3 percent. It does not affect the post-stratification weights, which are based only on education and age.

The primary estimates are unweighted and therefore average the experimental responses over the realized sample. To examine whether the results depend on its education and age composition, we also construct post-stratification weights for four cells formed by crossing college status with age at 30. Each weight is the cell's ACS population share divided by its share of the realized sample. Education and birth year were recorded for every respondent, so the weights require no imputation.

We chose this cross because each of its four cells has a population benchmark and is well represented in the sample. Both classifications are available for every respondent, so the weights can be constructed without imputation. Education also defines the experimental strata and the income grid, while age was added because the pilot revealed a substantial imbalance between the younger and older halves of the target population. We did not add further characteristics because the relevant background questions contain some missing responses and a finer cross would create small cells with substantial sampling variation. Table 2, Panel B, reports each cell's sample and population shares and its weight.

The 1st- and 99th-percentile caps did not alter any of the realized weights.

Age in the survey is the age a respondent reaches during 2026, calculated from her birth year, whereas the ACS benchmark records age when the ACS interview was completed. The weighted estimates are reported as robustness checks rather than replacing the unweighted primary results. Because the realized cell shares are close to their benchmarks, the weights range only from 0.948 to 1.102. Their design effect is 1.003, giving an effective sample size of 1,988 respondents.

## C Slutsky decomposition of the money price

A price change moves choice through two channels: a substitution effect, because the larger-family scenario is now relatively more expensive, and an income effect, because paying the higher price leaves the respondent with less household income. Separating them is the same exercise labor economists run on a wage change. The relevant welfare machinery here is compensated (Hicksian) choice probabilities in a random-utility model, following Small and Rosen (1981) and Dagsvik and Karlström (2005). Throughout, $\theta$ and $\mu$ denote the price and child-by-income coefficients from the child-by-income specification of equation (6), not the pooled coefficient of the main reduced-form demand equation; the two are numerically different because the pooled specification omits $C_{it}$ and averages across all three family-size contrasts and both directions of the price change.

For a given contrast with $k_f$ children in the smaller-family scenario and $k_m$ in the larger, holding the smaller scenario's income fixed and reducing the larger scenario's income by one log point identifies the uncompensated (Marshallian) response, $\theta - \mu k_m$; holding utility fixed at the smaller scenario's allocation instead identifies the compensated (Hicksian) response, $\theta - \mu k_f$. The compensated slope equals $\theta$ only when $k_f = 0$, since it depends on the base parity. Their difference is the income effect,

$$(\theta - \mu k_f) - (\theta - \mu k_m) = \mu \, \Delta k_{it}, \tag{13}$$

which is exact for the linear-probability coefficients estimated here; read as average marginal effects from a logit or probit it holds only approximately. The split is for the money price only: the time price admits an analogous decomposition, but the compensating adjustment there runs through time rather than income, so $\mu$ is not the relevant object and this derivation does not transfer.

Table A2 reports the decomposition contrast by contrast.

The zero-to-two contrast carries twice the income effect of either single-child contrast purely mechanically, because $\Delta k = 2$ rather than 1; this is an arithmetic consequence of the linear-in-count specification's single per-child income gradient $\mu$, not evidence that income matters more at that margin. Similarly, the zero-to-one and one-to-two contrasts share the same income effect, $\mu$, by construction. A specification with separate child-level dummies interacted with income would let the data speak to whether the first and second child differ in income sensitivity, but that is an exploratory extension beyond the registered linear specification. The finding to take from Table A2 is that substitution dominates the money-price response; it is not informative about how the income effect itself varies across family-size margins.

## D Structural estimates

Table A3 compares the income response implied by the structural model with the registered reduced-form estimate on a common probability scale. Because the conditional-logit coefficient is measured in latent-utility units, we calculate the change in predicted choice probability when both scenario incomes increase by 10 percent and average it across unequal-child tasks. The structural and reduced-form estimates are very similar.

Table A4 reports the conditional-logit estimates used to calculate the money and time equivalents in the main text. We estimate the model separately for college and non-college women. Because logit coefficients are measured relative to the scale of unobserved utility within each sample, individual coefficients should not be compared directly across the two columns. This scale factor cancels when the coefficients are combined to calculate the money and time equivalents.

Table 5 reports income equivalents at the median fielded income in each education stratum. Table A5 reports the corresponding equivalents at every fielded income level.

The same child valuations can be expressed in units of time rather than income. Table A6 reports the weekday-hour equivalents implied by the estimated paid-work coefficient.

## E The child penalty

Figure 3 reports willingness to pay net of the child penalty, defined as the annual value of the decline in a mother's earnings and the increase in her hours of paid and unpaid work that follow a birth. We evaluate the penalty at the household income of $125,000 used in the figure and construct a low and a high estimate. The penalties for a first and a second additional child are

$$\begin{aligned} \text{penalty}_1 &= p\, s\, Y + h_1\, v\, Y, \\ \text{penalty}_2 &= p\,(1-p)\, s\, Y + h_2\, v\, Y, \end{aligned}$$

where $Y$ denotes household income, $p$ the proportional reduction in a mother's earnings per child, $s$ her share of household earnings before a first birth, $h_1$ and $h_2$ the additional weekday hours of paid and unpaid work associated with a first and a second child, and $v$ the value of a weekday hour as a share of household income. The first term in each expression is the earnings penalty and the second is the time penalty.

For the earnings penalty we use the long-run US child penalty, the percentage by which mothers' earnings lie below their pre-birth trajectory five to ten years after a first birth. Kleven et al. (2019b) estimate it at 31 percent and Cortés and Pan (2023) at close to 40 percent. Because this penalty accumulates over all the children a mother has had by that point, and US mothers have on average 2.1 children ten years after a first birth (Kleven et al., 2019b), we interpret it as the penalty for two children and divide it equally between them. This gives a per-child earnings penalty of 15.5 percent in the low case and 20 percent in the high case. The equal division is consistent with Markussen and Strøm (2022), who find that first and second births reduce mothers' earnings in Norway by

similar amounts, 28 and 29 percent over the first three years.[11] The penalty for the second child is applied to earnings already reduced by the first, which gives the factor $1 - p$.

Since the earnings penalty is expressed relative to the mother's own earnings, whereas the scenarios specify household income, we require her share of household earnings. We use her share of couple earnings two years before a first birth, which matches the pre-birth baseline against which the penalty is measured. Musick et al. (2022) report that this share averaged 38 to 40 percent across recent cohorts, and was approximately 35 percent among women without a degree and 40 percent or more among graduates. We set it to 35 percent in the low case and 40 percent in the high case, a range that contains both the overall average and the share for any mixture of the two education groups.

We value a weekday hour at 2.35 percent of household income, the value of the respondent's time implied by the ratio of the time price to the money price (Table 3). Because the scenarios do not specify when within the ten-year horizon a child is born, the child may be anywhere from a newborn to about ten years old in the year the scenario describes. The high estimate therefore assumes a child under five, whose presence raises US women's weekday hours of paid and unpaid work by 1.9 hours (Craig and Mullan, 2010). The low estimate assumes a school-age child: we scale the regression estimate of 1.1 hours in Craig and Mullan (2010) by the ratio of mothers' time with a youngest child aged 5 to 11 to that with a youngest child aged 0 to 2 (0.5 to 1.4 in Australian data; Craig and Bittman, 2008), which yields 0.4 hours. A second child adds 64 percent of the first child's hours in the high case and 80 percent in the low case (Craig and Bittman, 2008; Bradbury, 2008). These hours are measured net of any reduction in paid work, whose effect on earnings is already captured by the earnings penalty.

The resulting penalty is \$7,900 a year for a first additional child and \$6,700 for a second in the low case, and \$15,600 and \$11,600 in the high case. The earnings penalty accounts for the larger share: \$6,800 of the low estimate and \$10,000 of the high estimate for a first additional child. The penalty is treated as known, so the confidence intervals on the net valuations do not reflect uncertainty in its inputs.

## F Experimental design

### *F.1 Attention check*

The instrument contains one attention check, identical in both versions. It sits among the background questions that follow the ten choice tasks, between relationship status and occupation, and reads:

> This is a question to check whether you are still paying attention and reading the questions carefully. Please select the two bottom options.

[11] We apply the same penalties whatever number of children a respondent already has, so the adjustment is identical across desired-fertility groups and across the two samples. For a mother, the additional children in the scenarios are higher-order births. The same study finds a somewhat smaller reduction after a third birth, 22 percent, so for respondents whose additional children would be third or later births our estimates may slightly overstate the penalty.

Five options follow as an agreement scale, from strongly agree to strongly disagree, with nothing stated to agree or disagree with, and respondents may select more than one. A response counts as correct only if it consists of the fourth and fifth options and nothing else. Because the check falls after the choice tasks, failing it cannot have affected the choices themselves; it defines one of the robustness samples in Section 4.6 rather than an exclusion. Appendix B.4 reports the pass rate.

### *F.2 Generation and assignment of experimental questions*

We generated a separate design for each education group, each containing 2,100 sets of ten questions and using the income levels calibrated for that group. To construct a profile, we drew independently and uniformly from the three additional-child levels, four income levels, and nine weekday allocations. We repeated this procedure for the second profile, redrawing it if it was identical to the first. The two profiles were otherwise treated symmetrically, so their contents were unrelated to their position on the screen.

Question sets were assigned sequentially without replacement within education group. This does not introduce a relationship between fielding order and task content because every set was generated independently under the same procedure. The 2,100 sets exceeded recruitment in both groups, so no set was reused.

Table A7 reports the properties of the realized design. The attribute levels are close to their intended shares, and the correlations confirm that the attributes were varied independently. The only substantial correlation is between paid-work and care and household hours ($r = -0.516$), which follows from restricting the design to the nine feasible weekday allocations.

## G Stated ideals and experimental choices

Tables A8 and A9 report the estimates underlying the comparison between stated ideals and contextualized demand in Section 4.5.

## H Robustness

We examine how the main estimates vary with sample composition, measures of response quality, progression through the choice tasks, and alternative specifications. We begin with the three confirmatory parameters: the money price $\theta$, the time price $\phi$, and the child-by-income interaction $\mu$. Table 9 in Section 4.6 re-estimates these parameters under alternative weights and sample restrictions while holding the specification fixed.

### *H.1 Alternative weights and analysis samples*

The analysis plan specified each of the weighting and alternative-sample analyses reported here.

Post-stratification weighting has almost no effect on the estimates. The money-price coefficient is $-0.2682$ in the primary estimates and $-0.2677$ after weighting; the corresponding time-price estimates are $-0.00629$ and $-0.00622$. The child-by-income interaction changes from 0.0284 to

0.0282. These small differences reflect the close correspondence between the realized education-by-age composition and the population cells used to construct the weights.

We also assess whether the estimates vary when we restrict the sample using measures of response quality and attention. We re-estimate the model after excluding respondents who failed the attention check. We also examine sensitivity to straight-lining by excluding respondents who selected the scenario displayed on the same side of the screen in all ten tasks. Both restrictions produce estimates close to those in the full sample. The estimates are also unchanged when we exclude the five observations flagged for conflicting eligibility information or possible repeat participation, as discussed in Appendix B.2.

The comprehension questions were designed to reinforce the instructions: respondents who answered incorrectly received corrective feedback before proceeding to the choice tasks. The pre-registered robustness analysis uses first-attempt responses and retains the 517 respondents (25.9 percent) who answered all three correctly. Because respondents entered the choice tasks having either answered correctly or received this feedback, we treat the first-attempt restriction as a supplementary measure of response quality. The money and time responses are larger in this sample, although the estimates are less precise given its smaller size.

We also use the time respondents spent reading the instructions and completing the ten choice tasks to construct two robustness samples. All respondents in the primary sample spent more than four minutes on the full survey, but total survey time includes the screening and background questions and any time for which the survey was left open. We therefore use two narrower measures: time on the ten choice tasks and time on the instructions and choice tasks together. Each check retains the slower two-thirds of respondents under the relevant measure. The money and time responses, as well as the child-by-income interaction, are larger in both timing samples. The similar pattern across the two measures suggests that noisier choices among faster respondents attenuate the full-sample price responses.

Perceived realism addresses a separate feature of the responses: whether the estimated trade-offs are similar among respondents who regarded the scenarios as plausible. This analysis retains the 1,336 respondents who reported that at least a majority of the scenarios felt realistic and produces estimates close to those in the full sample.

Across all the alternative samples, the estimates retain the downward response to the money and time prices and the positive income gradient.

### *H.2 Progression through the choice tasks*

Responses may change over a sequence of repeated choice tasks as respondents become more familiar with the choice format or as their attention changes. We therefore examine whether the effects of the money and time prices differ between tasks 1–5 and tasks 6–10. The money-price coefficient changes from $-0.253$ in the first five tasks to $-0.283$ in the final five, an increase of approximately 12 percent in magnitude. The corresponding change in the time-price coefficient is considerably larger: it increases in magnitude from $-0.00331$ to $-0.00920$. The differences are statistically significant

for both the money price ($p = 0.029$) and the time price ($p = 0.003$), and are jointly significant ($p = 0.0016$).

The particularly large change in the time response may reflect respondents becoming more familiar with the time-use profiles as they progress through the experiment. These profiles require respondents to consider paid work, care and household work, and the time remaining for leisure and sleep, and the first-attempt comprehension responses indicate that the time attribute was initially more difficult to understand than the income attribute. The results are therefore consistent with respondents placing greater weight on the time trade-off once they have gained experience with the choice format. The smaller increase in the money response suggests that there may also be some more general change in how respondents make their choices.

Stronger coefficients in the later tasks could reflect a general increase in choice consistency rather than a change in the relative importance of particular attributes. We examine this possibility by estimating the conditional-logit model separately for the first and second halves of the experiment and comparing these estimates with a model in which all coefficients can differ only by a common scale factor. The best-fitting common scale factor for the later tasks is 1.18. However, the restriction that the early and late coefficients differ only by this factor is rejected ($LR(6) = 23.32$, $p = 0.0007$). A general change in choice consistency therefore does not fully explain the pattern: the relative effects of the attributes also change, most notably through the substantially stronger response to time in the later tasks.

### *H.3 Income-focused choice patterns*

Some respondents selected the higher-income scenario in every task in which the two scenarios offered different incomes. This pattern could reflect a particularly strong preference for income, but it could also arise if respondents made their choices using the income attribute alone. We examine their other measures of engagement and their response to time when income cannot distinguish between the scenarios.

Among the 1,995 respondents who faced at least one task in which the scenarios differed in income, 351, or 17.6 percent, always selected the higher-income scenario. They faced a median of eight such tasks, and 90.3 percent faced at least six, so this pattern generally reflects repeated choices rather than a single comparison. Their response-quality measures are similar to those of other respondents. The attention-check pass rate is 95.4 percent, compared with 93.0 percent among other respondents, while 26.8 and 25.7 percent, respectively, answered all three comprehension questions correctly on their first attempt. Both groups spent a median of approximately 109 seconds on the ten choice tasks.

To determine whether these respondents considered attributes other than income, we examine tasks in which the two scenarios offer the same household income. Income cannot determine the choice in these tasks, allowing us to estimate the response to time while accounting for the difference in the number of children. Among the 331 income-focused respondents who faced at least one such task, an additional hour of paid work or care and household work reduces the probability of selecting

a scenario by 3.31 percentage points (SE 0.43 percentage points; $p < 0.001$). Applying the same specification to the equal-income tasks completed by other respondents gives a time response of 1.60 percentage points per hour. The response among income-focused respondents is therefore more than twice as large. Together with their similar performance on the response-quality measures, this indicates that they remain engaged with the time attribute despite consistently selecting higher income when it differs.

*H.4 Alternative specifications*

We examine whether the structural estimates are sensitive to key restrictions imposed by the main specification. We consider alternative ways of modeling income and use a mixed-logit model to allow preferences for additional children to vary across respondents. Together, these analyses assess whether the estimated valuations are stable across alternative functional forms and after accounting for unobserved preference heterogeneity.

The alternative income specifications preserve the main pattern of valuations. At the median income in each stratum, the preregistered linear-in-dollars specification produces estimates close to those from the primary log-income model. For non-college respondents, the income equivalents for the first and second additional child are \$4,700 and $-\$26,900$, compared with \$4,600 and $-\$26,300$ under log income. For college respondents, the corresponding estimates are \$24,900 and $-\$33,700$, compared with \$20,200 and $-\$29,300$ under log income (Table A10).

We next replace log income with indicators for each fielded income level, allowing the income response to be estimated flexibly. The log-income restriction is not rejected for non-college respondents ($p = 0.853$), but is rejected for college respondents ($p = 0.022$). The departure among college respondents is largely driven by the top of the distribution: choices respond less to the increase from \$240,000 to \$340,000 than the log specification would predict from the response at lower income levels. This weaker response to income produces larger income equivalents for the top interval (Table A11).

The fixed-coefficient model summarizes choices using common valuations of one and two additional children within each education stratum. If these valuations differ across respondents, imposing homogeneity may affect the estimates obtained from the nonlinear choice model. The ratio of coefficients can also depart from the valuation of a typical respondent if respondents who value children more are more or less responsive to income. We therefore estimate a mixed logit separately by education stratum in willingness-to-pay space, with each respondent's valuations, rather than her utility weights, as the parameters to be estimated:

$$V_{ijt} = \beta_{Yi}\Big[\ln Y_{jt} + \omega_{1i}\mathbf{1}\{k_{jt}=1\} + \omega_{2i}\mathbf{1}\{k_{jt}=2\} + \omega_W W_{jt} + \omega_H H_{jt} + \omega_\mu\, k_{jt}\,(\ln Y_{jt} - \overline{\ln Y}) + \omega_L L_{jt}\Big] + \varepsilon_{ijt},$$

where $\omega_{1i}$ and $\omega_{2i}$ are respondent $i$'s valuations of one and two additional children in log points of income, and $\ln \beta_{Yi}$, $\omega_{1i}$ and $\omega_{2i}$ are jointly normal with an unrestricted covariance matrix, so the income equivalent evaluated at the mean valuation is also the median. The model is estimated by

maximum simulated likelihood with 1,000 Halton draws per respondent, and standard errors are clustered by respondent.

Preferences vary substantially across respondents. A likelihood-ratio test rejects holding the income coefficient fixed in both strata ($\chi^2(3) = 445$ and 382, $p < 0.001$), and the income coefficient is negatively correlated with the valuation of a first additional child ($-0.17$ and $-0.24$): respondents who value a first child more respond less to income.

Allowing for this variation leaves the pattern of valuations unchanged and slightly increases their magnitude. For the first additional child, the median valuations are \$5,900 for non-college respondents (SE \$3,800) and \$28,600 for college respondents (SE \$6,100), compared with fixed-logit estimates of \$4,600 and \$20,200. For the second, they are $-\$32,800$ (SE \$4,800) and $-\$36,100$ (SE \$6,300), compared with $-\$26,300$ and $-\$29,300$. Each fixed-logit estimate lies inside the corresponding 95 percent confidence interval and is smaller in magnitude, so the fixed-logit valuations are, if anything, conservative estimates of the median.

### *H.5 Secondary results*

Table A12 extends the alternative weighting and sample restrictions to the secondary results. The hypotheses concern different outcomes and analysis samples, so the reported statistic differs across columns. In each case, we re-estimate the relevant specification using the alternative weights and samples described above. For H6, the table reports the main association of religiosity, desired additional fertility, and age 30 or older with baseline demand for children. Panel B reports the corresponding sample sizes.

The signs of the estimates for H4–H8 are unchanged across the alternative weights and sample restrictions. The ideals-asymmetry coefficient remains negative, while stated desire and each of the three measures of baseline demand remain positively associated with choices of larger families. The second-child shift and the full-time-work penalty remain negative. For H9, the restriction that the effects of care and household work are proportional to hours is not rejected in any of the samples, with $p$-values ranging from 0.313 to 0.978. There are some differences in magnitude, particularly in the smaller comprehension sample and under the timing restrictions, but they do not follow a common pattern across the secondary results.

Current and ideal numbers of children are recorded as six or more at the top of their respective response scales. We therefore repeat the two analyses that use desired additional fertility after excluding respondents whose current or ideal number of children is top-coded. The H5 coefficient increases from 0.079 to 0.088, while the H6 association between desired additional fertility and baseline demand increases from 0.079 to 0.090. The positive associations are therefore slightly stronger in the sample for which desired additional fertility can be calculated exactly.

Taken together, the weighting and sample restrictions preserve the directions of both the confirmatory and secondary results. The most systematic variation concerns the response to time: it is stronger among slower respondents and in the later choice tasks. The alternative income specifications and mixed-logit estimates also preserve the central ordering of the structural results,

although the flexible specification reveals a weaker income response at the top of the college income range. The main conclusions are therefore stable across the alternative samples and specifications, while the timing and task-progression analyses indicate that the magnitude of the time response is more sensitive to how respondents complete the experiment.

### *H.6 The second-child shift across money prices*

The H7 statistic measures the gap between the zero-to-one and one-to-two comparisons at a single point, where the two scenarios offer equal income. Figure A2 shows that the gap is present across the full range of fielded money prices.

## I Inference about unshown attributes

### *I.1 Open-text feedback*

At the end of the survey, respondents were asked: "Finally, thinking back to the scenarios that you were choosing between earlier in the survey, please describe briefly how you made your choices. This question helps us understand how people make these decisions. If you have any other comments about the survey, please include them here." Of the 1,995 respondents in the primary sample, 1,369 (68.6 percent) gave a nonblank response, with a median length of 16 words.

We screened the responses with keyword searches for each category below. The screened responses were then read and classified with the assistance of a large language model, using deliberately inclusive definitions that retained borderline cases, and every response classified as stating an assumption was checked by hand. A response is coded only if it states an assumption about something the scenarios did not show. A preference over a displayed attribute, such as a wish to work fewer hours, is not coded, and nor is a description of the respondent's own circumstances. The categories follow the second and third channels in Section 4.7. Forty-two respondents stated at least one such assumption: 24 about the source or control of household income, 16 about who supplies the remaining care, and seven about the job implied by the displayed income and paid-work hours. [12] No respondent inferred a partner's education, health or compatibility from the displayed household income, the characteristics that would make a higher income signal a more valued partner, and none inferred partner agreement or the stability of the relationship from the number of children shown.

**Source and control of household income.** Most of these responses concerned scenarios in which the respondent did no paid work, so that the household income would come from her partner, and where they referred to the partner it was to his earnings or working hours. About half treated that dependence as a drawback and favored scenarios in which they kept some paid work, even at a lower household income; several others said they would be content to stay at home if the partner earned enough. One inferred from a high income with no paid work of her own that her husband "would've been working himself to death." Another wrote that she initially chose the

[12] These counts sum to more than 42 because some responses fall in more than one category.

higher combined income but reconsidered when she realized that it could leave her "financially dependent on my partner." Others named the risks of that dependence directly: one wrote that she "would fear a divorce would make things difficult for me and the children," and another that relying on a partner's income alone "could lead to a potential for financial abuse." These responses are the basis for the statement in Section 4.7 that some respondents valued earnings of their own as protection should the relationship end.

**Who supplies the remaining care.** Every response in this category named either paid childcare or one of the two parents; none mentioned grandparents, other relatives or other informal care. Several treated a higher income as the means to buy childcare or other help, and others considered whether care would fall to the respondent or her partner when one of them was in paid work. Three read particular scenarios as the stay-at-home role: two of them identified that role with long hours of care and household work, which one described as feeling "closer to being a house-wife."

**The job behind the income and hours.** Four of the seven job-related responses judged whether the displayed income was plausible for the paid-work hours shown, for example that four hours of work at a six-figure income would not be practical, or read a high income with few hours as owning a business. One judged the displayed incomes against what her own occupation pays. The remaining two imagined a particular kind of job behind every scenario with paid work, one that was enjoyable or one that could accommodate a chronic illness, without linking it to the income shown. None inferred that a better-paid job would be more stressful or demanding.

**Other considerations.** Most responses describe an ordering of the displayed attributes, typically the number of children first, then income, then the division of the day, without stating any assumption about what the scenarios left unshown. A further 34 respondents discussed whether the displayed income was sufficient to support the number of children shown. One explained that "60k is not enough to raise two kids and myself," while another wrote that she could see her family having two children if she earned more than $300,000. We classify these affordability considerations as causal consequences of family size and income, the first channel in Section 4.7.

*I.2 The size of the inference channel*

Equation (12) writes the estimated money-price coefficient as the effect of the price alone plus an inference term. The inference term can be divided into two parts: the share $\pi$ of respondents who make the inference, and the slope $s$ of the inferred attribute on the money price among those who do. The proportional difference between the estimated coefficient and the effect of the price alone is then

$$\frac{\hat{\theta} - \theta_{\text{pure}}}{\theta_{\text{pure}}} \;=\; \pi \times m, \qquad m \equiv s\,\frac{\beta}{\theta_{\text{pure}}}, \tag{14}$$

where $m$, the strength of the inference, measures how far the inferred attribute moves an inferring respondent's choice for each unit of money price, relative to the effect of the price itself. Two features

of this expression follow directly. First, the difference is proportional to the share of respondents who make the inference, so a bound derived by assuming that every respondent does so falls in proportion when fewer do. Second, whenever the inferred attribute moves choices by no more than the price itself, so that $|m| \leq 1$, the proportional difference cannot exceed the share of respondents who make the inference.

**Job stress.** Suppose that respondents expect a better-paid job to be fast-paced rather than relaxed. Maestas et al. (2023) estimate that workers value a relaxed rather than a fast-paced job as highly as a 4.4 percent wage increase. Because the money price is measured in log income, a fast-paced job then affects choice as a money price 0.044 higher would, and equation (14) gives an understatement of the effect of the price alone equal to $0.044 \times \pi \times s$.

Job pace is binary, so between the lowest and highest incomes in the design the inferred pace can move at most once, from relaxed to fast-paced. If it rises in proportion to log income, the slope $s$ is the inverse of the range of log income: 1.13 in the non-college stratum ($60,000 to $185,000) and 0.96 in the college stratum ($130,000 to $340,000). If every respondent made the inference, the understatement would be $0.044/1.13 = 3.9$ percent in the non-college stratum and $0.044/0.96 = 4.6$ percent in the college stratum. If the inferred pace instead switches at a single income level, only choices between incomes on either side of that level are affected. Computing the slope across the pairs of income levels in each stratum, each equally likely under the design, the understatement ranges from 3.2 to 4.6 percent in the non-college stratum and from 4.0 to 5.5 percent in the college stratum, depending on where the switch occurs. Any belief that rises with income can be built from such switches, with weights summing to at most one, so 5.5 percent is the largest understatement that any such belief can produce when every respondent holds it. The bound is generous in a further respect: the 4.4 percent is a share of the worker's own wage, whereas the money price is measured in household income, of which a single job's earnings are only part.

Because the understatement is proportional to the share of respondents who make the inference, it lies well below 5.5 percent unless nearly every respondent pictured a more stressful job at higher incomes. If one respondent in ten did so, the understatement would be at most 0.55 percent. If the share equalled the share of respondents who raised the job behind the displayed income in their feedback at all, seven of 1,995, it would be at most 0.02 percent. None of those seven mentioned stress.

**Other job features.** Workers value several other job features, some more highly than a relaxed pace: Mas and Pallais (2017) estimate that workers would give up 8 percent of their wage for the option of working from home and 20 percent to avoid a schedule set by the employer at a week's notice, and Maestas et al. (2023) value a job without heavy physical activity at 12 to 15 percent of the wage. These features do not enter the bound because respondents would expect them to accompany higher pay rather than lower. Adrjan et al. (2026) find that higher posted wages lead workers to expect a more positive workplace culture, and that workers associate non-pecuniary amenities with higher pay; stress is the exception they document. An inference of this kind makes

the higher-income scenario more attractive than its income alone, which moves $\hat{\theta}$ away from zero and makes the willingness-to-pay estimates conservative.

*I.3 Partnered and unpartnered respondents*

No external estimate gives the value of an inferred partner relative to the income from which it is inferred, so we examine partner inference by comparing respondents who must imagine a partner with those who already have one. Every scenario describes a household with a partner. Respondents in a committed relationship, whether or not they live with their partner, can picture that partner whatever the income shown, and in the cognitive interviews, the pilot and the main study they indicated that they generally did so. Of the respondents who said in their feedback which partner they had in mind, 13 referred to an existing partner and six to a hypothetical one. Respondents who are not in a committed relationship must imagine a partner, and may infer a more valued one from a higher income. If they did, their money-price coefficient would be more negative than that of partnered respondents.

We estimate the reduced-form model with every coefficient interacted with an indicator for not being in a committed relationship, so that the interaction terms give the difference between the two groups. The money-price coefficient is −0.279 among the 1,265 partnered respondents and −0.251 among the 727 unpartnered respondents, a difference of 0.028 (standard error 0.019) in the opposite direction to the one partner inference predicts. The 95 percent confidence interval for the difference runs from −0.009 to 0.065, so it excludes a difference of more than 0.009 in the predicted direction, 3.3 percent of the partnered coefficient. If partnered respondents make no partner inference, this bounds the effect of the inference on the unpartnered coefficient. Because unpartnered respondents are 36.5 percent of the 1,992 respondents with a reported relationship status, the corresponding effect on the full-sample coefficient of −0.268 is at most about 1.2 percent.

## J External benchmarks for Section 5

*Stated ideals*

The parity-by-college standardization reweights the 2022–2023 NSFG sample of women aged 23–35 to match our sample's joint distribution across parity (0, 1, 2, 3+) and college attendance, then recomputes the RWANT share on the reweighted sample. The richer regression standardization instead regresses RWANT on age, age squared, parity, college attendance, partnership status and race in the NSFG, and predicts the NSFG share at our sample's covariate means. The trend extrapolation fits the two available NSFG data points (67.8% in 2017–2019, 57.4% in 2022–2023) as a linear trend in survey midpoint and projects it to our 2026 field date; because only two points are available, this is presented as a plausibility range, not a point estimate. On the secondary comparison, our 53.6% falls between the standardized NSFG shares who want another child (57.4%) and who expect another birth, ADDEXP (46.5%), nearer the wanting end of that interval — consistent with wanting and expecting being related but distinct constructs.

*The confirmatory hypotheses (H1, H2, H3)*

Table A13 constructs the money-price (H1) comparisons used in the main text.

Predicted values equal $\hat{\theta} \times$ % income, with $\hat{\theta} = -0.2682$ applied linearly, as in the main text, except for the two Poland rows (†), which use the mothers-only estimate $\hat{\theta}_{\text{mothers}} = -0.2721$ ($SE = 0.0135$) since Family 500+ recipients already have at least one child; the pooled estimate is statistically indistinguishable and predicts 2.6pp as well. The %-income figures already reflect each transfer's structure, so "recurring" and "one-off, annuitized" describe how that figure was built rather than a separate caveat on it: Spain's 0.77% annuitizes the one-off $3,900 payment over 18 years at 2.5% (factor 14.35, giving $272/year) against a $35,450 household-income base implied by González's reported 11%-of-median figure; Poland's 9.6% is instead the recurring $1,500/year (PLN 500/month) transfer over Bokun's reported median household income, used without adjustment. Converting González's and Gromadzki's aggregate national percentages to percentage points uses each country's total fertility rate divided by a 35-year childbearing window (Spain 1.4, Poland 1.3; both approximate), giving baseline annual birth probabilities of 4.0% and 3.7%; this baseline choice is the largest source of uncertainty here, since a higher-fertility subgroup baseline raises Poland's figure to about 0.6pp. The González estimates are the cumulative level sustained over the ∼30-month period the benefit was in force, though the authors note they cannot separate a shift in birth timing from a genuine increase in completed fertility; Gromadzki's estimate reflects only the immediate post-announcement jump, since Poland's transfer was never withdrawn within his sample.

Cesarini et al. (2023) estimate wealth effects on fertility from randomized lottery-prize size among Swedish players, separately by sex; we use only the female estimate, since our sample is entirely women. At the same ten-year horizon as our experiment, a 1,000,000 SEK ($150,000) win increases children born to women by 0.0078 ($SE = 0.0189$); the pooled and male estimates (0.0331 and 0.0558) are not used, since they reflect a larger, statistically distinguishable male response. Converting to a %-income basis follows Cesarini et al.'s own annuitization (20 years, 2% discount rate): a $100,000 prize equals a $5,996 permanent annual increase, or 22.4% of the $26,700 median individual disposable income in their sample. Applying $\hat{\mu} = 0.028$ (CI $[0.012, 0.045]$) to a shock this size predicts 0.0057 additional children (CI $[0.0024, 0.0091]$), against a rescaled actual of 0.0052 ($SE = 0.0126$) for women.

The income base above is individual, not household, disposable income; $\hat{\mu}$ is defined on household income, so this likely overstates the relevant % shock. This is not an inconsistency in Cesarini et al.'s own analysis — their internal benchmarking exercise (comparing annuitized lottery estimates to cross-sectional income gradients) uses the same individual-income definition throughout, so it is a difference in scope from our design, not an error on their part. Their own summary statistics report that 38% of the pooled estimation sample was married at the time of the lottery; for the remaining 62%, individual and household income roughly coincide, so the overstatement is confined to the married minority. Assuming, illustratively, that a married winner's household income is roughly double their individual income — comparable-earning spouses, an assumption

we cannot verify without their household-level data — a sample-weighted effective income base is $0.62 \times \$26{,}700 + 0.38 \times \$53{,}400 \approx \$36{,}850$, giving a revised shock of $\$5{,}996/\$36{,}850 \approx 16.3\%$ and a revised prediction of 0.0043 children. This sits just below the actual 0.0052, while the unadjusted 0.0057 sits just above it; because the doubling assumption is illustrative rather than estimated, we read this as showing the bias runs in a reassuring direction and is modest in size, not as a corrected estimate to prefer over the 0.0057 reported above.

Dahl et al. (2016) study Norway's extension of paid maternity leave from 18 to 35 weeks and find no significant effect on completed fertility. Andersson et al. (2006) document a positive but transitory "speed-premium" effect of Swedish parental-leave eligibility rules on birth spacing — a shift in timing, not a sustained change in family size. Neither offers a common-unit magnitude check against $\hat{\phi}$: the time reallocation in both cases is temporary and tied to the leave window around a single birth, not the decade-long, stable daily allocation our design describes.

*Religiosity*

Hayford and Morgan (2008) measure religiosity in the 2002 NSFG by how important religion is in a woman's daily life: very important, somewhat important, or not important. Women who report no religion are not asked the question, and Hayford and Morgan combine them with the not-important group. They compare period total fertility rates across the three groups, built from births to women aged 15 to 39 between January 1997 and December 2001. A total fertility rate is the number of children a woman would have if she experienced the period's birth rate at each age, so it does not depend on the age composition of each group. Women for whom religion was very important had a rate of 2.3, against 1.8 for women for whom it was not important (column 1 of Table A14).

We repeat their calculation on the 2022–2023 NSFG, with religiosity coded identically (`RELDLIFE`, with women reporting no religion in `RELIGION` combined into not important). The window is January 2017 to December 2021, the five calendar years before fieldwork began, as 1997–2001 was for the 2002 survey. The rate at each age divides live births at that age, taken from the pregnancy file with twins counted as two births, by the years women spent at that age during the window, and the total fertility rate sums these rates from 15 to 39. All estimates are weighted by `WGT2022_2023`. Hayford and Morgan report no standard errors; ours account for the NSFG's design, in which women are sampled in geographic clusters (`VECL`) within strata (`VEST`).

Column 2 shows that rates are lower in every group today, but the ordering is unchanged, and the gap between women for whom religion is very important and those for whom it is not is 0.70 children (standard error 0.15), at least as large as the half a child Hayford and Morgan found. The fall in levels need not mean smaller families. A period rate also falls when women have children later, because births postponed beyond the window, or beyond age 39, are not counted, and the average age at first birth in the United States has risen over this period.

Column 3 avoids this timing problem by counting the children ever born to women aged 40 to 49, whose childbearing is largely complete. The gap is 0.75 children (0.14), so the difference by religiosity is in how many children women have, not only in when they have them. In both surveys

religiosity is measured at interview, after most of the births being counted, so the table describes an association rather than an effect: having children may itself increase religious involvement.

## K Survey protocol

We reproduce below the full questionnaire, verbatim and in the order in which respondents saw it. The survey opened with a security check, a consent page and five screening questions, which determined eligibility (gender, birth year, US residence) and the version of the survey (education). Eligible respondents then completed the choice experiment, which consisted of instructions, a worked example, three comprehension questions, ten choice questions and a set of post-task questions. Figure A3 shows one of the choice questions as it appeared on screen. Respondents whose highest level of education was a four-year college degree or higher saw higher income levels than other respondents; the wording was otherwise identical, and both versions are shown where they differ.

*Consent*

> *[Security check (reCAPTCHA)]*
>
> Please complete the security check below to continue.
>
> *[New page]*
>
> **Survey on Life Preferences**
> **Evie Jamieson and Guy Michaels**
> **Department of Economics, London School of Economics**
>
> Thank you for your interest in this study, which explores the trade-offs people consider when making major life choices over the next decade.
>
> **What will my involvement be?**
> If you agree to take part, you will be asked to complete an anonymous online survey in which you will compare several hypothetical future scenarios and answer some general background questions. The survey should take approximately 10 minutes. Participation is voluntary, and you may skip questions you do not wish to answer, although answering some is necessary to complete the survey. We encourage you to take your time and reflect carefully as you work through each question.
>
> **Will I be compensated for taking part?**
> You will be compensated upon completing the survey. To ensure the quality of our research, we ask that you read each question carefully and take your time before responding. Compensation is contingent on completing the survey thoughtfully, so please do your best to engage genuinely with each question.
>
> **What will my information be used for?**
> Your responses will be used in our publications and future research, and, if you agree, may also be shared with other researchers through a data archive. Your participation will

be anonymous, meaning your name and personal details will not be used in any reports, publications or archived data resulting from the study. Your IP address will not be collected during this survey. Because your responses cannot be linked back to you, it will not be possible to withdraw your responses once you have completed and submitted the survey. The LSE Research Privacy Policy can be found here *[link in the survey]*. If you have any questions or concerns about this research, please contact the LSE research ethics managers at research.ethics@lse.ac.uk.

Q1. Please read and select the following statements if you agree with them:

☐ I have read and understood the explanations above
☐ I am 18 or over
☐ I agree to take part in the survey
☐ I give permission for the anonymised information I provide to be deposited in a data archive so that it may be used for future research

*[Shown if any statement in Q1 was not selected.]*

Thank you for taking the time to read about this study.

Participation is entirely voluntary, and you are under no obligation to take part.

If you did not mean to select this option, please use the back arrow to return to the consent form.

If you are sure you would like to exit, please click the forward arrow below.

*[Respondents who did not select all four statements were shown the message above and then left the survey.]*

*Screening questions*

Q2. When were you born? Please write in the year of your birth (use four digits for the year).
*[Drop-down list from 2025 to 1925]*

Q3. What is your gender?

Female; Male; Other (please specify): *[text box]*

Q4. What is your current employment status? (Please select all that apply).

Full-time employee; Part-time employee; Self-employed or business owner; Unemployed and looking for work; Full-time student; Part-time student; Not working and not looking for a job; On parental or maternity leave

Q5. Which category best describes your highest level of education?

Primary Education or Less; Some High School; High School Degree/GED; Some College; 2-year College Degree; 4-year College-Degree; Master's Degree (e.g., MA, MS, MBA);

Doctoral Degree (e.g., PhD, EdD, JD, MD)

Q6. Do you currently live in the United States?

Yes; No

*[Respondents born before 1990 or after 2003, who did not answer Female, or who did not live in the United States left the survey here. Respondents whose highest level of education was a four-year college degree or higher continued to the college version of the survey, and all others to the non-college version. The two versions differ only where shown.]*

*Instructions, worked example and comprehension questions*

On the following pages, you will see a series of questions. In each one, you will be presented with two possible versions of your life 10 years from now and asked to choose which you would prefer.

For all questions, please assume that you will have a partner or spouse 10 years from now, regardless of whether you currently have one. This assumption is applied to all respondents to ensure the scenarios are comparable and is not intended to reflect any expectation about your personal circumstances.

Each scenario describes three aspects of a possible future:

**1. Number of additional children**

This refers to any children you would have over the next 10 years, not counting any children you already have today. These additional children may be biological (including via surrogacy), step-, or adopted.

**2. Combined income of you and your partner**

This shows the combined annual income you and your partner (or spouse) would have before taxes, 10 years from now. This income level would be reached gradually over this period before stabilizing. For example, a figure of $100,000 means you and your partner would together be earning $100,000 per year before taxes 10 years from now. Please ignore inflation (you may assume that prices do not change over this period).

**3. Typical weekday**

This describes how you would usually spend a typical weekday (Monday to Friday) 10 years from now, divided into three categories:

A. Paid work (including self-employment and commuting time, if applicable)

B. Care and household tasks (housework, errands, and care for children and elderly or dependent adults)

C. Leisure & sleep (the remaining time)

These hours add up to 24.

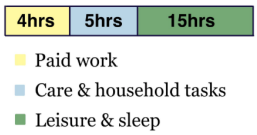
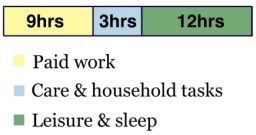
For each question, compare the two scenarios and select the one you would personally

prefer, even if neither is ideal. There are no right or wrong answers. We are only interested in what you prefer.

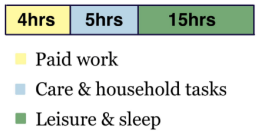
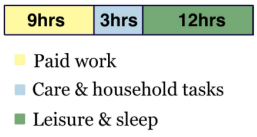
*[New page]*

The example below shows what each question will look like. In every question, you will compare two possible versions of your life 10 years from now.

Each scenario describes three aspects of a possible future: the number of additional children, the combined income with your partner, and how a typical weekday is divided across paid work, care and household tasks, and leisure and sleep. The three weekday categories always sum to 24 hours.

| | **Scenario 1** | **Scenario 2** |
|---|---|---|
| Number of additional children | One child | No children |
| Combined income of you and your partner (before tax) | $175,000 | $240,000 |
| Typical weekday |  |  |

*[College version]*

| | **Scenario 1** | **Scenario 2** |
|---|---|---|
| Number of additional children | One child | No children |
| Combined income of you and your partner (before tax) | $95,000 | $140,000 |
| Typical weekday |  |  |

*[Non-college version]*

In the example above, Scenario 1 includes one additional child, a combined income of [*college version:* $175,000; *non-college version:* $95,000], and a weekday split of 4 hours of paid work, 5 hours of care and household tasks, and 15 hours of leisure and sleep. Scenario 2 includes no additional children, a combined income of [*college version:* $240,000; *non-college version:* $140,000], and a weekday split of 9 hours of paid work, 3 hours of care and household tasks, and 12 hours of leisure and sleep.

For each question, please select the scenario you personally prefer, even if neither feels ideal.

*[New page]*

Q7. Imagine a woman who has 1 child today. If a scenario shows 1 additional child, how many children would she have in total 10 years from now?

Only the child she has today; The child she has today, and 1 more child; The child she has today, and 2 more children

*[Shown on the same page once "The child she has today, and 1 more child" is selected in Q7.]*

That's correct. The number of additional children shows how many children would join your family over the next 10 years. Every scenario also includes any children you already have.

*[Shown on the same page once "Only the child she has today" is selected or "The child she has today, and 2 more children" is selected in Q7.]*

That's incorrect. She would have 2 children in total. The number of additional children shows how many children would join your family over the next 10 years. Every scenario also includes any children you already have.

*[New page]*

Q8. In the scenarios, "combined income of you and your partner" means:

The household's total yearly income before paying taxes; The household's yearly income after paying taxes; The amount the household has left after taxes and household spending

*[Shown on the same page once "The household's total yearly income before paying taxes" is selected in Q8.]*

That's correct. In these scenarios, income means the household's total yearly income before paying taxes or household spending.

*[Shown on the same page once "The household's yearly income after paying taxes" is selected or "The amount the household has left after taxes and household spending" is selected in Q8.]*

That's incorrect. In these scenarios, income means the household's total yearly income before paying taxes or household spending.

*[New page]*

Q9. In a scenario where the typical weekday bar shows 0 hours of paid work, this means:

I am not doing any paid work, but my partner may still be working; I am not doing any paid work, and my partner is also not doing any paid work; Neither I nor my partner are doing any paid work, and the bar shows our hours combined

*[Shown on the same page once "I am not doing any paid work, but my partner may still be working" is selected in Q9.]*

That's correct. The bar describes your own typical weekday only. Your partner's schedule is not shown and may differ from yours.

*[Shown on the same page once "I am not doing any paid work, and my partner is also not doing any paid work" is selected or "Neither I nor my partner are doing any paid work, and the bar shows our hours combined" is selected in Q9.]*

That's incorrect. The bar describes your own typical weekday only, not your partner's. Your partner's schedule is not shown and may differ from yours.

*Choice questions*

Q10–19. *[Ten choice questions, each on its own page. Each shows two scenarios in a table, as in Figure A3. The levels in each question were drawn from the experimental design.]*

($k$/10) Please choose your preferred scenario in 10 years time:

| | Scenario 1 | Scenario 2 |
|---|---|---|
| Number of additional children | *[level]* | *[level]* |
| Combined income of you and your partner (before tax) | *[level]* | *[level]* |
| Typical weekday | *[level]* | *[level]* |
| | ◯ | ◯ |

*Post-task questions*

Q20. How many of the scenarios that you were choosing between in the previous section felt realistic to you?

All or almost all; A majority, but not almost all; A minority, but more than almost none; Almost none

Q21. How many children do you have? Please include biological, step-, surrogate and adopted children.

0; 1; 2; 3; 4; 5; 6+

Q22. In total, including any children you already have, how many children would you ideally like to have?

0; 1; 2; 3; 4; 5; 6+

Q23. Which of the following best describes your current relationship status?

Married or in a civil union/domestic partnership; In a committed relationship and living together; In a committed relationship but not living together; Not currently in a committed relationship

Q24. This is a question to check whether you are still paying attention and reading the questions carefully. Please select the two bottom options.

Strongly agree; Somewhat agree; Neither disagree nor agree; Somewhat disagree; Strongly disagree *[more than one answer allowed]*

*[New page]*

Q25. Which of the following categories best describes your occupation?

Farmer or agricultural laborer, rancher, fisher; Manual laborer (e.g., factory worker, miner); Tradesperson (e.g., mechanic, welder, painter, plumber, tailor); Service worker (e.g., driver, waiter, cook, retail worker, cashier, barber, janitor, housekeeper); Clerical worker (e.g., secretary, bookkeeper, receptionist); White-collar worker (e.g., manager, executive, businessperson, salesperson, accountant, banker); Professional (e.g., doctor, lawyer, engineer, IT/computer programmer); Medical or social worker (e.g., nurse, EMT, pharmacist, social worker); Protective service worker (e.g., police, fire); Educational service worker (e.g., teacher, professor); Public servant (e.g., bureaucrat, politician, military); Homemaker / stay-at-home parent; Self-employed / small business owner (excluding farm owners); Other (please specify) *[text box]*; Don't know

*[New page]*

Q26. What is your present religion, if any?

Protestant (for example, Baptist, Methodist, Non-denominational, Lutheran, Presbyterian, Pentecostal, Episcopalian, Reformed, Church of Christ, etc.); Roman Catholic; Church of Jesus Christ of Latter-day Saints (Mormon / LDS); Orthodox (such as Greek, Russian, or some other Orthodox church); Jewish; Muslim; Buddhist; Hindu; Atheist or agnostic; No religion; Other (please specify): *[text box]*; Prefer not to answer

Q27. How important is religion in your life?

Very important; Rather important; Not very important; Not at all important

Q28. Which of the following best describes where you live?

In a city or town (of 50,000 or more people), or a suburb or area that is part of one; In a rural area, or a small town that is not part of a larger city or town

Q29. What is your race and/or ethnicity? *(Select all that apply)*

American Indian or Alaska Native; Asian; Black or African American; Hispanic or Latino; Middle Eastern or North African; Native Hawaiian or Pacific Islander; White; Other (please specify) *[text box]*; Prefer not to answer

*[New page]*

Q30. In the past 12 months, what were your total earnings from work, before taxes? *Include wages, tips, bonuses, and self-employment income.*

$0–$9,999; $10,000–$34,999; $35,000–$59,999; $60,000+; Don't know; Prefer not to answer

*[New page]*

Q31. Finally, thinking back to the scenarios that you were choosing between earlier in the survey, please describe briefly how you made your choices. This question helps us understand how people make these decisions.

If you have any other comments about the survey, please include them here.

*[Text box]*

To keep your response anonymous, please avoid names or other specific personal details.

*[Respondents whose total time on the survey was four minutes or less were returned to the panel as having completed too quickly; all others were returned as complete.]*

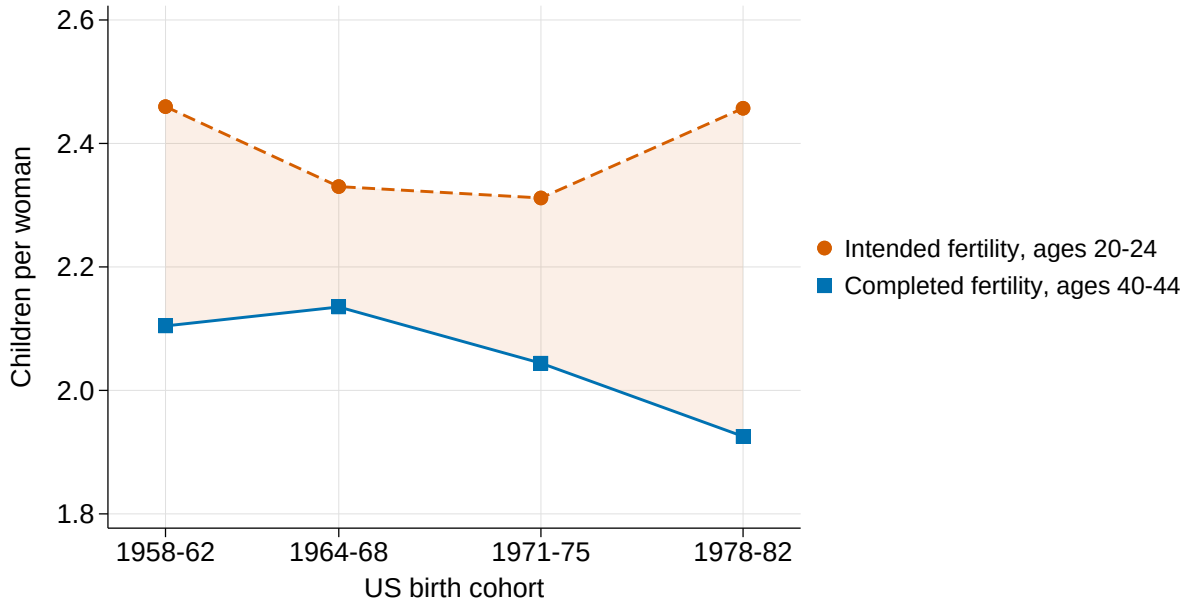



Figure A1: Fertility intentions and completed fertility, by US birth cohort

*Notes:* Intended fertility is mean total intended parity—children already born plus additional births expected—among women aged 20–24. Completed fertility is mean children born among women aged 40–44. Each pair uses separate cross-sections restricted to the same approximate birth-year band. Estimates use each cycle's final survey weight. The average shortfall is the unweighted mean of the four cohort-specific gaps.

*Source:* Authors' calculations from National Survey of Family Growth public-use files: 1982, 1988, 1995, 2002, 2006–2010, 2015–2017, and 2022–2023.

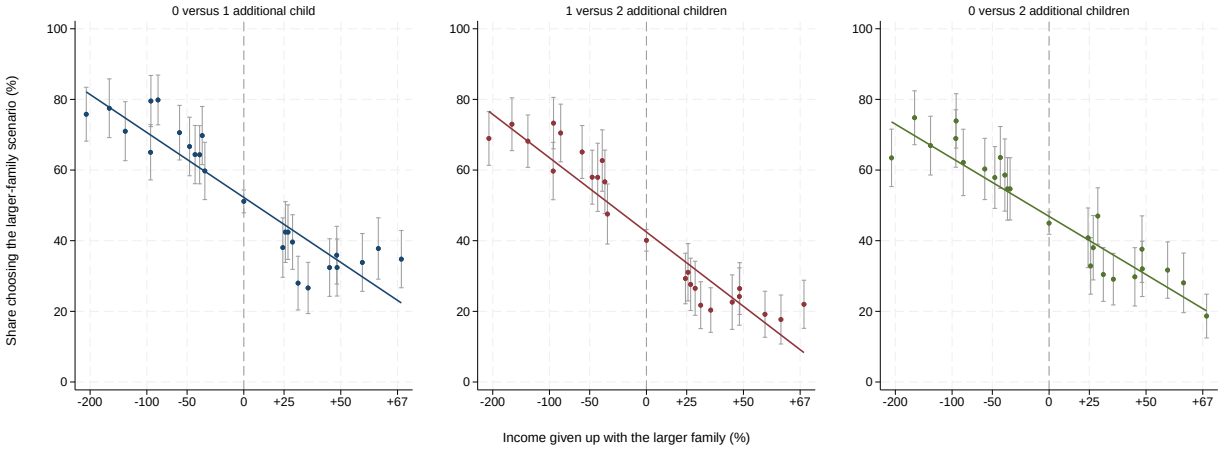



Figure A2: Choice shares and fitted money-price relationship, by family-size comparison

*Notes:* Points are the share choosing the larger-family scenario at each fielded money price, with 95 percent confidence intervals clustered on the respondent. Lines are fitted values from Equation (11), which gives each comparison its own price slope. Respondent fixed effects absorb the intercept, so the height of the lines is a normalization: the zero-to-one line passes through the mean share and mean price of its tasks, and the distances between the lines are estimates.

**(1/10) Please choose your preferred scenario in 10 years time:**

| | **Scenario 1** | **Scenario 2** |
|---|---|---|
| Number of additional children | One child | Two children |
| Combined income of you and your partner (before tax) | $240,000 | $130,000 |
| Typical weekday | 4hrs \| 3hrs \| 17hrs<br>Paid work<br>Care & household tasks<br>Leisure & sleep | 9hrs \| 5hrs \| 10hrs<br>Paid work<br>Care & household tasks<br>Leisure & sleep |
| | ○ | ○ |

Figure A3: Example choice question

*Notes:* The first of the ten choice questions, as displayed in the college version of the survey. Respondents in the non-college version saw the same layout with their own income levels.

Table A1: Survey completion times, primary sample

| Timing measure | Observations | 10th percentile | 25th percentile | Median | 75th percentile | 90th percentile |
|---|---|---|---|---|---|---|
| Total survey time | 1,995 | 280 | 336 | 446 | 690 | 1,252 |
| Ten choice tasks | 1,994 | 59 | 78 | 109 | 168 | 277 |
| Instructions and choice tasks | 1,994 | 78 | 105 | 155 | 250 | 438 |

*Notes:* All times are in seconds. Page-level timing was unavailable for one respondent, so the two measures of time on the experimental material have 1,994 observations.

Table A2: Contrast-specific Slutsky decomposition of the money-price slope

| Contrast | Hicksian ($\theta - \mu k_f$) | Marshallian ($\theta - \mu k_m$) | Income effect ($\mu \Delta k$) |
|---|---|---|---|
| Zero to one (extensive) | −0.2392 | −0.2677 | 0.0284 |
| | (0.0124) | (0.0092) | (0.0084) |
| Zero to two (extensive) | −0.2392 | −0.2961 | 0.0569 |
| | (0.0124) | (0.0126) | (0.0169) |
| One to two (intensive) | −0.2677 | −0.2961 | 0.0284 |
| | (0.0092) | (0.0126) | (0.0084) |

*Notes:* Both coefficients are taken from the child-by-income specification of equation (6), estimated on the primary sample with standard errors clustered on the respondent. Each cell combines them according to the contrast's starting and ending family size, and its standard error is obtained by the delta method from their joint covariance. In the first two rows the smaller scenario has no children, so $k_f = 0$ and the compensated slope is $\theta$.

Table A3: Income interaction on the probability scale

| | Reduced form (registered test) | Structural (registered counterpart) | Tasks |
|---|---|---|---|
| All unequal-child tasks | 0.36 | 0.34 | 13,442 |
| | [0.15, 0.57] | [0.15, 0.53] | |
| One-child difference | 0.27 | 0.25 | 9,052 |
| | [0.11, 0.43] | [0.11, 0.40] | |
| Two-child difference | 0.54 | 0.51 | 4,390 |
| | [0.23, 0.86] | [0.23, 0.79] | |

*Notes:* Entries are percentage-point changes in the probability of choosing the larger-family scenario following a 10 percent increase in both scenario incomes; brackets contain 95 percent confidence intervals. The reduced form is estimated on the 13,442 unequal-child tasks. The structural model is estimated on all 19,950 tasks, with probability changes averaged over the unequal-child tasks shown in each row. Structural intervals use 2,000 Krinsky–Robb draws from the respondent-clustered covariance matrix.

Table A4: Conditional-logit estimates underlying willingness to pay

| | Non-college | College |
|---|---|---|
| One additional child | 0.101 | 0.194 |
| | (0.050) | (0.056) |
| Two additional children | −0.234 | −0.094 |
| | (0.068) | (0.078) |
| Two versus one, incremental | −0.335 | −0.288 |
| | (0.047) | (0.051) |
| Paid-work hours | −0.041 | −0.056 |
| | (0.006) | (0.007) |
| Care and household hours | −0.042 | −0.055 |
| | (0.009) | (0.009) |
| Log household income | 1.360 | 1.609 |
| | (0.070) | (0.084) |
| Additional children × log income | 0.066 | 0.083 |
| | (0.046) | (0.055) |

*Notes:* Each column reports a separate conditional-logit estimation of Equation (7) on the primary tasks of one education stratum, and these coefficients underlie the willingness-to-pay estimates in Table 5. Standard errors, clustered on the respondent, are in parentheses.

Table A5: Willingness to pay for additional children, at every fielded income level

| Stratum | Income ($000s) | Additional children | | |
|---|---|---|---|---|
| | | 0 to 1 | 1 to 2 | 0 to 2 |
| Non-college | 60 | $1,705 | −$18,208 | −$16,083 |
| | | [−2,613, 6,023] | [−23,281, −13,136] | [−23,791, −8,374] |
| | 95 | $4,643 | −$26,337 | −$20,662 |
| | | [−1,263, 10,550] | [−33,696, −18,977] | [−30,236, −11,089] |
| | 140 | $9,212 | −$35,769 | −$24,698 |
| | | [392, 18,033] | [−47,319, −24,219] | [−39,709, −9,687] |
| | 185 | $14,390 | −$44,419 | −$27,333 |
| | | [1,760, 27,020] | [−61,263, −27,574] | [−50,405, −4,262] |
| College | 130 | $13,294 | −$23,883 | −$8,842 |
| | | [5,277, 21,312] | [−32,397, −15,368] | [−21,614, 3,931] |
| | 175 | $20,162 | −$29,301 | −$6,797 |
| | | [10,333, 29,990] | [−39,880, −18,722] | [−20,939, 7,344] |
| | 240 | $30,902 | −$36,092 | −$2,092 |
| | | [17,550, 44,254] | [−51,017, −21,167] | [−20,951, 16,766] |
| | 340 | $48,776 | −$44,835 | $7,987 |
| | | [28,055, 69,498] | [−69,049, −20,622] | [−23,807, 39,781] |

*Notes:* Each entry is the change in annual household income that leaves a respondent indifferent between the two family sizes, evaluated at each income level shown to her education group. Estimates come from the conditional logits of Table 5. A positive figure is income she would give up for the larger family, and a negative figure is the compensation that family must provide. Because income enters in logs, the same share of income is worth more dollars at higher incomes, so dollar valuations rise with income even when the share is unchanged. Figure 2 reports the valuations as shares of income. Brackets contain 95 percent confidence intervals calculated by the delta method, with standard errors clustered on the respondent.

Table A6: Willingness to pay for additional children, in weekday hours

| Additional children | Non-college | College |
|---|---|---|
| One versus none | 2.5 | 3.5 |
| | [−0.0, 5.0] | [1.3, 5.7] |
| Two versus one | −8.2 | −5.1 |
| | [−11.6, −4.9] | [−7.2, −3.0] |
| Two versus none | −5.8 | −1.7 |
| | [−9.5, −2.1] | [−4.4, 1.0] |

*Notes:* From the conditional logits of Table 5. A positive entry is the number of extra weekday hours of paid work, taken from leisure and sleep, that a respondent would accept to have the larger family. A negative entry is the number of weekday hours of paid work the larger family would have to remove to leave her as well off. Measuring the hours in care and household work instead gives near-identical figures. Brackets contain 95 percent confidence intervals (delta method, standard errors clustered on the respondent).

Table A7: Properties of the realized design

| | | Correlation between attributes | | | |
|---|---|---|---|---|---|
| Attribute and levels | Share of options (%) | (1) | (2) | (3) | (4) |
| (1) Additional children: 0, 1, 2 | 33.0 / 33.4 / 33.5 | 1 | | | |
| (2) Household income: four levels | 24.4 / 25.2 / 25.0 / 25.4 | −0.008 | 1 | | |
| (3) Paid hours: 0, 4, 9 | 22.1 / 44.3 / 33.5 | 0.003 | 0.000 | 1 | |
| (4) Care hours: 2, 3, 5, 8 | 22.1 / 22.3 / 33.6 / 21.9 | 0.003 | −0.003 | −0.516 | 1 |
| Tasks holding the number of children equal: 32.6 per cent | | | | | |

*Notes:* Primary sample, comprising 1,995 respondents, 19,950 tasks and 39,900 scenario options. Shares are of scenario options and sum to 100 within each attribute; they are a check that the randomization ran as written rather than a property of the design. Correlations use log income for the income attribute; the same correlations computed on the within-task differences, which are what identify the coefficients, agree to within 0.01. Income levels are \$60,000, \$95,000, \$140,000 and \$185,000 for non-college women and \$130,000, \$175,000, \$240,000 and \$340,000 for college women. All 108 profiles in the three-by-four-by-nine factorial were fielded in both strata, and every respondent completed all ten tasks. Paid and care hours are correlated because only nine of the twelve paid-by-care combinations were fielded and leisure and sleep is the residual of the day.

Table A8: Registered H4 among childless women

| | Childless women |
|---|---|
| *Probability of choosing the stated number of children* | |
| Stated ideal of zero | 0.820 |
| | (0.014) |
| Stated ideal of one | 0.497 |
| | (0.033) |
| Registered log-odds contrast | −1.531 |
| | (0.162) |
| Tasks | 1,311 |
| Respondents | 542 |

*Notes:* The table reports registered exploratory H4 among childless respondents whose stated ideal is zero or one, using tasks that offer a choice between no children and one child. Let $y_{it} = 1$ if respondent $i$ chooses one child in task $t$, and let $A_i = \mathbf{1}\{a_i = 1\}$, where $a_i$ is stated desired additional fertility. The registered saturated logit is

$$\log\left(\frac{\Pr(y_{it}=1)}{1-\Pr(y_{it}=1)}\right) = \gamma_0 + \gamma_1 A_i.$$

The first two rows convert the fitted log odds into the probability of choosing the stated number: $1 - \Lambda(\gamma_0)$ when $a_i = 0$ and $\Lambda(\gamma_0 + \gamma_1)$ when $a_i = 1$. The registered contrast is the difference between the corresponding log odds of adherence, $2\gamma_0 + \gamma_1$. Standard errors in parentheses are clustered on the respondent.

Table A9: Desired additional fertility and experimental choices

| | Larger-family choice |
|---|---|
| Desired additional fertility | 0.079 |
| | (0.006) |
| Money price | −0.270 |
| | (0.009) |
| Time price | −0.0074 |
| | (0.0011) |
| Position | 0.012 |
| | (0.008) |
| Constant | 0.393 |
| | (0.010) |
| Model | LPM |
| Design-stratum fixed effects | Yes |
| Tasks | 13,319 |
| Respondents | 1,976 |

*Notes:* Standard errors in parentheses are clustered on the respondent. The linear probability model uses unequal-child tasks and asks whether stated desire predicts choice, with design-stratum fixed effects $\delta_{c(i)}$:

$$y_{it} = \delta_{c(i)} + \psi a_i + \theta P_{it} + \phi T_{it} + \lambda L_{it} + u_{it}.$$

Table A10: Income equivalents under the two money scales, at the stratum median income

| | Non-college, at \$95,000 | | College, at \$175,000 | |
|---|---|---|---|---|
| | Log income (primary) | Linear in dollars (secondary) | Log income (primary) | Linear in dollars (secondary) |
| 0 to 1 | \$4,643 | \$4,694 | \$20,162 | \$24,936 |
| | [−1,263, 10,550] | [−2,454, 11,916] | [10,333, 29,990] | [10,742, 39,099] |
| 1 to 2 | −\$26,337 | −\$26,923 | −\$29,301 | −\$33,694 |
| | [−33,696, −18,977] | [−33,955, −20,398] | [−39,880, −18,722] | [−44,880, −22,719] |
| 0 to 2 | −\$20,662 | −\$22,478 | −\$6,797 | −\$10,528 |
| | [−30,236, −11,089] | [−31,928, −13,991] | [−20,939, 7,344] | [−26,573, 4,997] |

*Notes:* Dollars, with 95 percent confidence intervals in brackets. Within each education group, the log-income column repeats the estimates of Table 5. The linear-in-dollars column re-estimates the same model with income entering linearly, in units of \$10,000, so utility carries $\beta_Y Y + \mu\, k\, (Y - \bar{Y})$ and the equivalent solves

$$Y_k^*(Y) = \frac{\beta_Y Y - \alpha_k + k\,\mu\,\bar{Y}}{\beta_Y + k\,\mu},$$

with the figure reported being $Y - Y_k^*(Y)$ converted back to dollars. Under the linear scale the valuation would not vary with income at all if $\mu$ were zero, which is why it is reported at the median rather than across the grid.

Table A11: Valuations from the income-dummy specification

| Stratum | Income step (\$000s) | 0 to 1 | 1 to 2 | 0 to 2 |
|---|---|---|---|---|
| Non-college | 60–95 | \$2,945 | −\$18,985 | −\$16,168 |
| | | [−2,147, 8,037] | [−23,865, −14,105] | [−22,946, −9,389] |
| | 95–140 | \$6,883 | −\$27,439 | −\$20,861 |
| | | [−746, 14,512] | [−35,412, −19,465] | [−30,958, −10,764] |
| | 140–185 | \$12,631 | −\$37,360 | −\$25,314 |
| | | [471, 24,791] | [−51,780, −22,940] | [−43,041, −7,587] |
| College | 130–175 | \$17,439 | −\$24,485 | −\$7,840 |
| | | [7,111, 27,767] | [−33,544, −15,426] | [−20,595, 4,916] |
| | 175–240 | \$23,248 | −\$26,625 | −\$4,304 |
| | | [11,706, 34,790] | [−36,402, −16,849] | [−17,900, 9,293] |
| | 240–340 | \$51,619 | −\$46,528 | \$2,221 |
| | | [26,539, 76,699] | [−67,869, −25,188] | [−27,244, 31,687] |

*Notes:* The specification of Table 5 with $\beta_Y \ln Y_{jt}$ replaced by a free dummy for each fielded income level,

$$V_{jt} = \alpha_1 \mathbf{1}\{k_{jt}=1\} + \alpha_2 \mathbf{1}\{k_{jt}=2\} + \beta_W W_{jt} + \beta_H H_{jt} + \sum_{r=2}^{4} \eta_r \mathbf{1}\{Y_{jt} = Y_r\} + \mu\, k_{jt}\,(\ln Y_{jt} - \overline{\ln Y}) + \lambda L_{jt} + \varepsilon_{jt},$$

so each gap between adjacent levels has its own price of income, $\eta_{r+1} - \eta_r$, and the valuation for a step uses the local price from that gap rather than a common slope. 95 percent confidence intervals in brackets. This nests the log form, and a test of that restriction is not rejected for non-college women ($p = 0.853$) and is rejected for college women ($p = 0.022$).

Table A12: Robustness of the secondary results

| | H4 | H5 | H6: baseline demand | | | H7 | H8 | H9 |
|---|---|---|---|---|---|---|---|---|
| Analysis sample | Ideals asymmetry | Stated desire | Religiosity | Desired fertility | Age 30+ | Second-child shift | Full-time penalty | Care linearity |
| Primary | −1.531 | 0.079 | 0.063 | 0.079 | 0.070 | −0.098 | −0.081 | F=0.53 |
| | (0.162) | (0.006) | (0.006) | (0.006) | (0.013) | (0.011) | (0.014) | (p=0.591) |
| Post-stratification weighted | −1.514 | 0.079 | 0.063 | 0.078 | 0.070 | −0.098 | −0.081 | F=0.51 |
| | (0.162) | (0.006) | (0.006) | (0.006) | (0.013) | (0.011) | (0.014) | (p=0.598) |
| Passed attention check | −1.713 | 0.088 | 0.061 | 0.086 | 0.065 | −0.100 | −0.088 | F=0.86 |
| | (0.173) | (0.006) | (0.006) | (0.006) | (0.013) | (0.011) | (0.014) | (p=0.425) |
| Passed all comprehension checks | −1.697 | 0.104 | 0.071 | 0.104 | 0.107 | −0.076 | −0.104 | F=0.02 |
| | (0.315) | (0.011) | (0.011) | (0.011) | (0.025) | (0.021) | (0.026) | (p=0.978) |
| Not straight-lined | −1.540 | 0.082 | 0.064 | 0.082 | 0.072 | −0.099 | −0.080 | F=0.43 |
| | (0.163) | (0.006) | (0.006) | (0.006) | (0.013) | (0.011) | (0.014) | (p=0.648) |
| Slower two terciles by choice-task time | −1.352 | 0.079 | 0.060 | 0.079 | 0.065 | −0.105 | −0.114 | F=0.84 |
| | (0.195) | (0.007) | (0.007) | (0.007) | (0.016) | (0.013) | (0.017) | (p=0.430) |
| Slower two terciles by instructions and tasks | −1.393 | 0.084 | 0.059 | 0.084 | 0.064 | −0.093 | −0.101 | F=0.78 |
| | (0.201) | (0.007) | (0.007) | (0.007) | (0.016) | (0.013) | (0.017) | (p=0.458) |
| Found scenarios realistic | −1.372 | 0.061 | 0.070 | 0.064 | 0.069 | −0.112 | −0.064 | F=1.16 |
| | (0.205) | (0.007) | (0.007) | (0.007) | (0.016) | (0.013) | (0.017) | (p=0.313) |
| Excluding five anomalous records | −1.531 | 0.080 | 0.062 | 0.080 | 0.069 | −0.098 | −0.081 | F=0.51 |
| | (0.162) | (0.006) | (0.006) | (0.006) | (0.013) | (0.011) | (0.014) | (p=0.601) |
| *Panel B. Sample sizes (number of respondents)* | | | | | | | | |
| Analysis sample | H4 | H5 | | H6 | | | H7–H9 | |
| Primary | 542 | 1,976 | | 1,941 | | | 1,995 | |
| Post-stratification weighted | 542 | 1,976 | | 1,941 | | | 1,995 | |
| Passed attention check | 502 | 1,851 | | 1,817 | | | 1,864 | |
| Passed all comprehension checks | 138 | 511 | | 508 | | | 517 | |
| Not straight-lined | 539 | 1,953 | | 1,918 | | | 1,971 | |
| Slower two terciles by choice-task time | 350 | 1,308 | | 1,288 | | | 1,319 | |
| Slower two terciles by instructions and tasks | 353 | 1,316 | | 1,293 | | | 1,329 | |
| Found scenarios realistic | 299 | 1,320 | | 1,293 | | | 1,336 | |
| Excluding five anomalous records | 542 | 1,971 | | 1,936 | | | 1,990 | |

*Notes:* Each row applies the indicated weighting or sample restriction to the specification defining the corresponding result. Entries for H4, H5, H6, H7 and H8 are coefficients, with respondent-clustered standard errors in parentheses. The three H6 columns report the main associations of religiosity, desired additional fertility and age 30 or over with choosing the larger-family scenario when measured money and time costs are equal. These are cross-sectional associations estimated with education-stratum fixed effects. H9 reports the joint $F$-statistic for the care-linearity restrictions and its $p$-value. The estimation samples differ across hypotheses: H4 is restricted to the relevant childless respondents and zero-versus-one tasks; H5 requires non-missing desired additional fertility; H6 requires non-missing values for all three respondent characteristics; and H7–H9 use all applicable choice tasks. Panel B therefore reports respondent counts separately by hypothesis.

Table A13: Constructing the H1 (money-price) comparisons, ordered by external observation window, longest to shortest

| Study | % income | Predicted | Reported | Horizon |
|---|---|---|---|---|
| González (2013) | 0.77% (one-off, annuitized) | 0.21pp | 6% agg. → 0.24pp | ∼30 mo sustained |
| González & Trommlerová (2023) | 0.77% (one-off, annuitized) | 0.21pp | 3% agg. → 0.12pp | ∼30 mo sustained |
| Bokun (2024) | 9.6% (recurring) | 2.6pp† | 1.5pp | 1 yr, direct |
| Gromadzki (2024) | 9.6% (recurring) | 2.6pp† | 6% agg. → 0.22pp | ∼9 mo, announcement only |

Table A14: Fertility by importance of religion, 2002 and 2022–2023 NSFG

| | Hayford & Morgan Total fertility rate, 1997–2001 | NSFG 2022–23 Total fertility rate, 2017–2021 | NSFG 2022–23 Children ever born, ages 40–49 |
|---|---|---|---|
| Very important | 2.3 | 2.12 | 2.29 |
| | | (0.13) | (0.10) |
| | | N = 1,438 | N = 536 |
| Somewhat important | 2.1 | 1.67 | 1.88 |
| | | (0.11) | (0.10) |
| | | N = 1,222 | N = 380 |
| Not important/no religion | 1.8 | 1.41 | 1.54 |
| | | (0.08) | (0.10) |
| | | N = 2,159 | N = 557 |
| Gap (very − not important) | 0.5 | 0.70 | 0.75 |
| | | (0.15) | (0.14) |

*Notes:* Column 1 is from Hayford and Morgan (2008), Table 2. Columns 2 and 3 use the 2022–2023 NSFG, weighted by `WGT2022_2023`. Column 2 is the period total fertility rate for ages 15 to 39 over January 2017 to December 2021; column 3 is mean children ever born (`PARITY`) among women aged 40 to 49. Religiosity is the importance of religion in daily life, with women reporting no religion combined into not important. Standard errors in parentheses account for the survey's strata and clusters; those in column 2 are computed by the delta method. Birth month is not on the public file, so in computing exposure for column 2 each woman's birth is placed at the midpoint of the twelve months her age at interview allows. N is the number of women contributing to each estimate.