Online Appendix to "Not Yet or Never? Economic Constraints, Not Lifestyle Preferences, Are the Main Perceived Barrier for Adults Still Open to Parenthood"

Contents

ONLINE APPENDIX A. BACKGROUND INFORMATION QUESTIONNAIRE

Respondents complete the background information questionnaire in either English or Korean. For brevity, we report only the English version here. Country of residence is not elicited in the questionnaire because it is verified through Pureprofile's panel records, which restrict participation to respondents residing in Singapore, South Korea, or the United States. Among U.S. respondents, we further restrict participation to those of East Asian ethnicity.

1. **Please select the age group you belong to.**

   Please select one response.

   - ○ Under 21
   - ○ 21–30 y.o
   - ○ 31–40 y.o
   - ○ 41–50 y.o
   - ○ Over 50 y.o
   - ○ Prefer not to say

2. **Please select your gender.**

   Please select one response.

   - ○ Male
   - ○ Female

3. **Please provide us with an estimate of your monthly gross income.**

   Please select one response.

   - ○ Under USD 2,300
   - ○ Between USD 2,300 and USD 4,600
   - ○ Over USD 4,600
   - ○ Prefer not to say

4. **Please select the highest level of education you have achieved.**

   Please select one response.

   - ○ Less than High School
   - ○ High School
   - ○ University
   - ○ Postgraduate
   - ○ None of the above

5. **Which best describes your current employment status?**

   Please select one response.

   - ○ Employed full-time
   - ○ Employed part-time
   - ○ Self-employed
   - ○ Student

- ◯ Homemaker
- ◯ Currently not employed, seeking work
- ◯ Currently not employed, not seeking work
- ◯ Prefer not to say

6. **Which of the following best describes the industry you currently work in (or most recently worked in)?**

   Please select one response.

- ◯ Education / Academia
- ◯ Healthcare / Medical Services
- ◯ Social Work / Psychology / Counseling
- ◯ Information Technology / Software / AI
- ◯ Finance / Banking / Insurance
- ◯ Consulting / Business Services
- ◯ Public Sector / Government / Policy
- ◯ Legal / Law Enforcement / Judiciary
- ◯ Media / Journalism / Communications
- ◯ Arts / Entertainment / Design
- ◯ Retail / Hospitality / Food Services
- ◯ Manufacturing / Engineering / Construction
- ◯ Transportation / Logistics / Travel
- ◯ Science / Research / Biotechnology
- ◯ Energy / Utilities / Environmental Services
- ◯ Agriculture / Forestry / Fisheries
- ◯ Nonprofit / NGO / Community Organization
- ◯ Freelance / Self-Employed
- ◯ Student
- ◯ Homemaker / Not currently employed
- ◯ Prefer not to say

7. **Which best describes your current relationship status?**

   Please select one response.

- ◯ Single
- ◯ In a relationship
- ◯ Married / Partnered
- ◯ Divorced / Separated
- ◯ Widowed
- ◯ Prefer not to say

8. **How many children do you currently have?**

   Please select one response.

- ◯ I don't have children
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ More than 3

For Question 3, "Please provide us with an estimate of your monthly gross income," we use country-specific income brackets to account for differences in income levels across the three countries. For Singapore, the brackets are "Under SGD 3,000," "Between SGD 3,000 and SGD 6,000," and "Over SGD 6,000." For South Korea, the brackets are "Under KRW 3,200,000," "Between KRW 3,200,000 and KRW 6,400,000," and "Over KRW 6,400,000."

ONLINE APPENDIX B. PROMPTING THE AI INTERVIEWER

We prompt GPT to conduct a one-on-one interview in the persona of a sociologist asking how individuals reason about whether to have children.[1] Our prompt follows the interviewer design developed in Chopra and Haaland (2023), Geiecke and Jaravel (2024), and Hwang et al. (2025), and combines state-of-the-art techniques such as role-play, structured reasoning, few-shot examples, instruction following, self-consistency checks, and an agentic control structure. We refer the reader to those papers for a detailed discussion of the general methodology of AI-led interviewing and its connection to current best-practice prompting techniques. Here we summarize the components that are specific to our context.

1. **Persona and Role.** The interviewer is cast as a sociologist trained in qualitative, non-directive, trauma-informed interviewing, with a warm, empathetic, and non-judgmental demeanor. Its communication style follows the OARS framework (Open Questions, Affirmations, Reflections, and Summaries), which is widely used in qualitative interviewing (Miller and Rollnick, 2012). The interview is conducted in English or Korean, depending on the language the respondent chooses at the start of the session.

2. **Core Objective.** The interviewer is told that the goal of the interview is to elicit each respondent's own narrative and sense-making about whether to have children. It is also given the broader motivation of the study: to understand the considerations that shape contemporary childbearing decisions in settings where fertility is below replacement, such as Singapore, South Korea, and the United States.

3. **Knowledge Base (Internal Reference).** The prompt also includes concise summaries of seven leading theories of fertility behavior, which serve as an internal reference for the model: (i) economic cost–benefit calculations (Becker, 1960); (ii) the second demographic transition and the shift towards post-materialist values (Lesthaeghe, 2010); (iii) life-course ideals and prerequisites (Thornton, 2005); (iv) gender roles and the "double burden" of career and motherhood (Myong et al., 2021); (v) future uncertainty and risk aversion (Kreyenfeld, 2010); (vi) social spillovers and competitive parenting (Kim et al., 2024); and (vii) shifting social

[1] We use GPT-5 (gpt-5-2025-08-07) with reasoning effort set to low to balance response time against interview quality.

norms and peer influence (Balbo and Barban, 2014). These theories are offered as examples of factors that may bear on childbearing decisions, not as a script: the interviewer is instructed to remain strictly non-leading and never to reveal them to respondents. Embedding a balanced set of perspectives also mitigates the risk that the model overweights frameworks that are heavily represented in its pre-training corpus while underemphasizing others that are theoretically important but less frequently discussed.

4. **Interview Protocol.** Once the respondent's background and context have been established, the interviewer follows a step-by-step protocol:

   (a) Begin with an opening question, common to all respondents, that invites them to describe their current thinking about having children.

   (b) Probe with approximately ten open-ended, non-directive follow-up questions that focus on why and how respondents hold their views, including the beliefs, emotions, trade-offs, and experiences behind them.

   (c) Ask a fixed final-sweep question ("Is there anything else that feels important to your thoughts or feelings that we haven't touched on yet?") and, if new topics emerge, follow up with two to three additional probes.

   (d) Ask a multiple-choice question about how often the respondent thinks about having children (1 = Never, 2 = Sometimes, 3 = Often, 4 = Always).

   (e) Provide a neutral summary of the conversation in no more than 150 words and ask the respondent to evaluate it along four dimensions: the fidelity of the summary, the overall quality of the interview, their preference between an AI and a human interviewer, and their preference between a conversation and an essay format.

   (f) Conclude with a brief thank-you statement.

5. **Guiding Principles.** The prompt also lays out the methodological and conversational standards the interviewer must follow: remaining non-directive and non-leading, encouraging detailed narratives, avoiding repetition, asking only one question per message, and maintaining a compassionate and natural conversational flow. As in Hwang et al. (2025), the prompt also includes pacing strategies. The interviewer is instructed to treat each question as a scarce

opportunity within a budget of approximately ten follow-ups, so that the core objective of the interview is achieved within a reasonable conversation length..

6. **Self-Correction Checklist.** Before generating each question, the interviewer applies a self-consistency checklist. It evaluates whether the proposed question satisfies a set of criteria and revises the question if any criterion is not met. The criteria include fit with the objective, strategic value within the question budget, forward-looking framing, clarity, variety in style and pattern, formatting, non-directiveness, and final polish.

7. **Exception Handling.** Finally, the prompt includes instructions for handling prompt-injection attempts and inappropriate or problematic content. Because childbearing can touch on sensitive personal topics, the interviewer is also instructed to refrain from offering medical or legal advice and to redirect such requests to qualified professionals while continuing the interview. These instructions help keep the model on task, maintain appropriate professional boundaries, and preserve the integrity of the interview.

In the final sample, interviews last 13.9 minutes on average (median: 10.8 minutes), measured as the time elapsed between the first and last message of a session.

Online Appendix B.1 Hard-coded questions in the AI-led interview

To keep the interview consistent across respondents and to avoid biases that could arise from variation in question phrasing, six elements of the interview are hard-coded into the prompt: the opening question and the five multiple-choice questions (MCQ1–MCQ5) described in the protocol above. All hard-coded questions are translated into Korean by hand to ensure equivalence across languages; only the English versions are reproduced below.

Online Appendix B.1.1 Opening question

The opening question is asked at the start of every interview.

Opening question

Hi, I appreciate you taking the time to talk with me. This interview is about understanding your perspectives on having children. To start, could you tell me how you personally think and feel about the idea of having children in your life?

ONLINE APPENDIX B.1.2 MULTIPLE-CHOICE QUESTIONS: MCQ1–MCQ5

After the open-ended portion of the interview, the interviewer administers five hard-coded multiple-choice questions in a fixed order. MCQ1 asks how often the respondent thinks about having children; MCQ2 asks the respondent to evaluate the interviewer's summary of the conversation; and MCQ3–MCQ5 capture the respondent's overall evaluation of the AI-led interview experience.

MCQ1: Frequency of thought about having children

Just one quick follow-up question—How often do you find yourself thinking about having children?

1 (Never)
2 (Sometimes)
3 (Often)
4 (Always)

Please only reply with the associated number.

MCQ2: Summary fidelity

How well does the above summarize how you think about whether to have children?

1 (It's a poor summary)
2 (It's a fair summary)
3 (It's a good summary)
4 (It's a very good summary)

Please only reply with the associated number.

MCQ3: Overall interview quality

To wrap up, just three quick multiple-choice questions about your experience with this AI-led interview. First, how would you rate the overall quality of the interview?

1 (Poor)
2 (Fair)
3 (Good)
4 (Very good)

Please only reply with the associated number.

MCQ4: Preferred interviewer

Second, in the future, would you prefer to be interviewed by

1 (an AI-led interviewer)
2 (a human interviewer)
3 (I am indifferent)

Please only reply with the associated number.

MCQ5: Preferred format

Third, which format do you prefer? An AI-led conversation (like here) or writing out your thoughts as a 'short essay' in response to an open-ended question?

1 (I prefer AI-led conversations)
2 (I prefer writing out my thoughts to an open-ended question)
3 (I am indifferent)

Please only reply with the associated number.

Online Appendix C. Prompts for the post-interview analysis

We construct eight prompts corresponding to the three steps of the post-interview analysis described in Section 2 of the main draft.[2] Step 1, which constructs argument maps, uses three prompts; Step 2, which identifies recurrent narratives in a data-driven manner, uses one; and Step 3, which validates, refines, consolidates, and assesses the resulting narratives, uses four. This section summarizes the key elements of these prompts. Online Appendix C.4 describes an additional prompt, separate from the argument-map pipeline, that classifies the deliberation state that each transcript reaches; this classification is also used in Section 2 of the main draft. The prompting techniques are similar to those described in Online Appendix B and Hwang et al. (2025), and for brevity we omit explanations of components that overlap with earlier material.

Online Appendix C.1 Step 1: Constructing argument maps

This step converts each interview transcript into an argument map, that is, a directed acyclic graph (DAG) whose nodes are the considerations the respondent raises and whose directed edges record how the respondent connects these considerations to one another and, ultimately, to having or not having children. We use the terms argument map and DAG interchangeably below. The step consists of three prompts. The first prompt generates a DAG for each transcript. Because the 1,074 DAGs are generated independently, the same underlying construct can appear under different node names in different DAGs. The second prompt therefore consolidates the nodes generated across all transcripts into a standardized set, and the third prompt maps each original node to a node in that set. The output of Step 1 is a master node dictionary, which lists the standardized nodes, and 1,074 DAGs whose node names are standardized across transcripts. Both serve as inputs to Step 2.

Online Appendix C.1.1 Generating a DAG for each transcript

The key elements of the prompt used to generate a DAG for each transcript are as follows:

1. **Core Objective.** We task the model with constructing a compact structural causal model

[2] We use GPT-5.2 (gpt-5.2-2025-12-11), the most recent GPT model available at the time of analysis. The only exception is residual-pattern extraction during iterative validation, for which we use GPT-5-mini (gpt-5-mini-2025-08-07) to balance cost and performance. Reasoning effort ranges from medium to xhigh across steps, while model and reasoning effort are held fixed within each task and step.

(Pearl, 2009) and the corresponding DAG that explain the respondent's articulated reasoning about whether to have children. The emphasis is on the causal mechanisms underlying that reasoning, as the respondent perceives them, rather than on correlations, associations, or proxy indicators.

2. **Input.** The model receives the respondent's interview transcript, truncated immediately before the closing multiple-choice questions (Online Appendix B.1.2), so the DAG draws only on the open-ended portion of the interview. All nodes must be grounded in what the respondent says during the interview; the interviewer's questions are used only for context and clarification.

3. **Method.** The prompt provides the following modeling instructions:

   (a) *Target specification:* The only terminal nodes allowed are the two outcome nodes, *HavingChildren* and *NotHavingChildren*. Both may appear in the same DAG when the respondent expresses mixed views.

   (b) *Node construction:* Each node receives a name and a definition. It must be supported by verbatim transcript snippets and must reflect a single underlying construct. When the DAG is generated, we also instruct the model to assign a domain tag to each node. The tag indicates the type of consideration the node represents and takes one of six values: economic, psychological, intergenerational, societal, biological, or other. For example, nodes related to income or housing are assigned to the economic domain. The domain tag provides a node-level classification that is comparable across all 1,074 DAGs and thereby helps the model standardize nodes that were generated independently in different DAGs (see Online Appendix C.1.2).

   (c) *Edge construction:* Edges represent direct causal effects from one node to another, as perceived by the respondent. We require each edge to have clear support in the transcript. When uncertain, the model is instructed to omit the edge.

   (d) *Graph constraints:* The graph must be acyclic, and its size should be proportional to the complexity of the transcript, typically between 5 and 25 nodes.

4. **Evidence and Endorsement Rules.** Concepts introduced by the interviewer may enter the DAG only if the respondent explicitly endorses them, and every supporting snippet must be a

verbatim quote from the respondent. These rules guard against attributing to the respondent ideas that the interviewer raised.

ONLINE APPENDIX C.1.2 STANDARDIZING NODES

The key elements of the prompt used to standardize nodes are as follows:

1. **Core Objective.** We task the model with consolidating the nodes from the individual DAGs into a set of standardized nodes, merging only nodes that represent the same underlying construct.

2. **Input.** The prompt is run separately for each node domain. The model receives the names and definitions of the nodes in that domain, generated in the previous step (Online Appendix C.1.1).

3. **Method.** Within a domain, two nodes are merged only if they match on four dimensions: underlying construct, polarity, level of abstraction, and source. For example, support from a partner is not merged with government support for families. Each standardized node is assigned a concise name derived from the names or definitions of its constituent original nodes. Catch-all buckets, such as a generic "Emotion" or "Money," are disallowed unless the constituent original nodes are themselves equally broad. The outcome nodes *HavingChildren* and *NotHavingChildren* are protected and are never folded into standardized nodes. Every original node must map to exactly one standardized node; if a node matches none, the model creates a singleton node for it.

ONLINE APPENDIX C.1.3 MAPPING ORIGINAL NODES TO STANDARDIZED NODES

The key elements of the prompt used to map original nodes to the standardized set are as follows:

1. **Core Objective.** We task the model with mapping each original node to exactly one standardized node.

2. **Input.** The model receives the names and definitions of the original nodes, together with the set of standardized nodes produced by the previous prompt.

3. **Method.** The mapping is based on construct identity and uses both the alignment of definitions and lexical similarity.

ONLINE APPENDIX C.2 STEP 2: DATA-DRIVEN IDENTIFICATION OF RECURRENT NARRATIVES

The key elements of the prompt used to identify recurrent narratives are as follows:

1. **Core Objective.** We task the model with extracting the narratives in the 1,074 DAGs, that is, recurring patterns of reasoning across respondents. Each narrative captures a coherent causal path, with a clear driver, through which respondents connect their considerations to having or not having children. The model identifies these narratives by looking for paths that recur across respondents' DAGs.

2. **Input.** The prompt is run separately for each of the two outcomes, *HavingChildren* and *NotHavingChildren*. In each run, the model receives the part of each respondent's standardized DAG that leads to that outcome, together with the node dictionary from Step 1.

3. **Method.** The prompt is organized around five guiding principles:

   (a) *Grounding:* The model uses only the input DAGs and the node dictionary. It does not introduce external concepts or unsupported mechanisms.

   (b) *Merging equivalent narratives:* Candidate narratives are merged when they stem from the same source, such as the self, the job, or society, and work through the same underlying mechanism, even when they are labeled differently.

   (c) *Root causes:* Each narrative is defined by what drives the respondent's reasoning, not by the state that results. When a node only describes a state, such as not feeling ready, the model looks further back in the DAG for its cause, such as housing costs. Patterns that consist only of an unexplained desire or aversion, without an identifiable driver, are discarded.

   (d) *Individual versus systemic constraints:* The model keeps three kinds of constraints apart: a respondent's own shortage of money, time, or energy; external systems and their costs, such as the housing market or the childcare system; and cultural or institutional pressure, such as workplace culture or gender-role expectations. This keeps individual constraints from being confused with systemic ones.

(e) *Parsimony and coverage:* The model keeps the number of narratives as small as possible while still covering the patterns that recur across respondents. A pattern must appear in the DAGs of roughly ten or more distinct respondents to qualify; patterns found in only a few respondents are rejected.

4. **Output per Narrative.** Each narrative is summarized by (i) a short title, in the form of a noun phrase of one to three words; (ii) a description of no more than 50 words that connects the driver to the outcome of having or not having children; (iii) a core path, that is, the most common path behind the narrative, written with the node names used in the DAGs; and (iv) path variants, that is, other recurring paths that reflect the same reasoning. The model must use node names exactly as they appear in the DAGs, so that each narrative can be traced back to them.

Online Appendix C.3 Step 3: Validation, refinement, consolidation, and assessment

This step consists of four prompts. The first prompt measures how much of each respondent's reasoning the current narratives leave unexplained and records the unexplained paths. The second prompt refines the narrative catalog based on these results. These two prompts are applied in alternating rounds. The third prompt combines similar narratives into a smaller final catalog. The fourth prompt scores each respondent's DAG against the final catalog on salience and direction.

Online Appendix C.3.1 Validation

The key elements of the prompt used for validation are as follows:

1. **Core Objective.** We task the model with assigning each respondent a residual score, which measures how much of the respondent's reasoning about having children the current narrative catalog leaves unexplained, and with recording the unexplained reasoning.

2. **Input.** The model receives three inputs: (i) the current narrative catalog; (ii) an individual respondent's DAG; and (iii) the corresponding interview transcript, truncated immediately before the closing multiple-choice questions (Online Appendix B.1.2). The transcript serves only as context for interpreting the DAG.

3. **Method.** The model proceeds in four steps:

   (a) *Read the respondent's DAG.* The model reads the DAG and identifies all paths that lead to *HavingChildren* or *NotHavingChildren*.

   (b) *Identify what is already explained.* The model sets aside the nodes and edges that an existing narrative already covers. A narrative covers its topic in both directions: if the catalog contains a narrative about a given factor, reasoning in which that factor pushes towards having children and reasoning in which it pushes against are both treated as explained. Only reasoning with a distinct causal logic counts as unexplained.

   (c) *Assign a residual score.* The model assigns a residual score from 1 to 4. A score of 4 indicates that the respondent's main driver is missing from the catalog; 3 indicates that an important secondary argument is unexplained; 2 indicates that only minor details are missing; and 1 indicates that the catalog covers all of the respondent's reasoning.

   (d) *Record unexplained patterns.* The model records the unexplained paths, together with a short description, as residual patterns. The residual patterns of respondents with a residual score of 3 or 4 are passed on to refinement.

ONLINE APPENDIX C.3.2 REFINEMENT

The key elements of the prompt used to refine the narrative catalog are as follows:

1. **Core Objective.** We task the model with refining the narrative catalog, using the residual patterns identified during validation.

2. **Input.** The model receives two inputs: (i) the current narrative catalog, including each narrative's description, core path, and path variants; and (ii) the residual patterns aggregated across respondents during validation.

3. **Method.** The model refines the catalog in three ways. First, it merges redundant narratives, including mirror-image pairs: a reason for having children and the corresponding reason against are treated as two sides of a single narrative. Second, it carefully broadens the descriptions of existing narratives so that they absorb the residual patterns, while keeping each

narrative's core causal structure. Third, if substantial residual patterns remain, it proposes at most one new narrative, based on the largest group of residual patterns.

Online Appendix C.3.3 Consolidation

The key elements of the prompt used for consolidation are as follows:

1. **Core Objective.** We task the model with combining similar narratives in the refined catalog into a smaller final catalog, without losing the reasoning that the original narratives capture.
2. **Input.** The model receives the refined narrative catalog produced in the previous step.
3. **Method.** Narratives that share the same underlying mechanism are combined into a broader narrative. Each broader narrative (i) keeps all core paths and path variants of the narratives it combines, (ii) records which narratives it combines, and (iii) receives a newly written description that reflects all of them rather than only one. The model must also preserve direction: if all the combined narratives lead only to not having children, the broader narrative cannot describe a route to having children, and vice versa.

Online Appendix C.3.4 Assessment

The key elements of the prompt used to score each final narrative for each respondent are as follows:

1. **Core Objective.** We task the model with assigning two scores for each respondent and each final narrative: a *salience* score, which captures how central the narrative is to the respondent's reasoning, and a *direction* score, which captures whether the narrative pushes the respondent towards or against having children.
2. **Input.** The model receives three inputs: (i) the final narrative catalog; (ii) the respondent's DAG; and (iii) the respondent's interview transcript, truncated immediately before the closing multiple-choice questions (Online Appendix B.1.2).
3. **Method.** The two scores draw on different sources. Salience is read primarily from the DAG, based on how central the narrative's path is in the respondent's reasoning. Direction is read primarily from the transcript, based on whether the respondent describes the narrative

as pushing towards or against having children. Narratives are scored independently, so a respondent can receive a high salience score on several narratives.

4. **Salience rubric (1–4).** A score of 4 indicates that the narrative is clear and dominant in the respondent's DAG; 3 indicates a moderate contribution that is clearly present but not primary; 2 indicates a weak and fragmented presence; and 1 indicates absence.

5. **Direction rubric ($\{-1, 0, +1\}$).** A score of +1 indicates that the narrative pushes the respondent towards having children, -1 indicates that it pushes the respondent against having children, and 0 indicates that it is not applicable. By construction, salience= 1 implies direction= 0.

Online Appendix C.4 Classifying the deliberation state of each transcript

Separately from the argument-map pipeline, we classify the deliberation state each respondent's reasoning reaches (Section 2 of the main draft), using GPT-5.2 with high reasoning effort. The prompt's key elements are as follows:

1. **Core Objective.** We task the model with determining the deliberation state the respondent's reasoning reaches about having children, taking into account every consideration the respondent raises—financial, relational, health-related, or value-based. The instructions emphasize that the object is the current status of the decision as expressed in the respondent's reasoning, not the respondent's personality, sentiment towards children, or predicted future behavior.

2. **Input.** The model receives the interview transcript truncated immediately before the closing deliberation-frequency question, so the classification is blind to the respondent's self-reported deliberation frequency, the AI-generated summary, and the interview-evaluation answers. Only interviewee statements carry evidential weight; interviewer questions are context.

3. **Classes.** Exactly one of the four classes is assigned: *active consideration* (actively considering and inclined towards having children, with no unmet condition holding the decision back), *postponement* (keeps children open but delays until conditions are met), *ambivalent* (genuinely unresolved about ever having children), and *permanent refusal* (settled against having children, whatever the grounds). Postponement and refusal are distinguished not by the reasons for saying no, but by whether the person names conditions under which they would consider

having children. Leaving that possibility open indicates postponement; ruling it out indicates refusal, even when the decision is attributed to circumstances.

4. **Output.** One assigned class, an ordinal confidence grade (high/medium/low), a verbatim interviewee quote of at most 25 words grounding the label, and a rationale of at most 25 words.

Figure 2 reports the validation of the resulting classification against the withheld deliberation-frequency question.

ONLINE APPENDIX D. REWEIGHTING TO THE POPULATION OF CHILDLESS ADULTS

Our samples are recruited through an online panel using quotas on gender, age, income, and education and are therefore not nationally representative. This section assesses the implications of this sampling design by estimating the share of childless adults in each country who would be classified as *not yet*. To extrapolate from our samples to these populations, we use post-stratification, reweighting each country sample to match the age and gender composition of its childless adult population. We also reweight by gender and education and, for the United States, by gender and income. Table 5 reports the results.

ONLINE APPENDIX D.1 PROCEDURE

Within each country, we divide respondents into eight cells defined by gender and four age groups: 21–30, 31–40, 41–50, and over 50. For each cell, we calculate the share of respondents classified as *not yet* using our transcript-based classification (Online Appendix C.4).

We next use external population statistics to estimate the number of childless adults in each cell of the general population. We convert these counts into each cell's share of the country's childless adult population, yielding weights that sum to one.[3] The reweighted *not-yet* share is the weighted average of the eight cell-specific *not-yet* shares, using these population weights. We report two versions of this calculation. The first uses all eight cells and targets childless adults aged 21–79 in the United States, 20–79 in South Korea, and 20 and older in Singapore. The second restricts the target populations to ages 21–49 in the United States and 20–49 in Singapore and South Korea. Panel C of Table 5 reports the cell shares for both target populations. Finally, the row "Three countries, equal weight" reports the simple average of the three country-specific estimates.

In separate exercises, we reweight by gender and education and, for the United States, by gender and income. The cells distinguish respondents with and without a bachelor's degree or respondents in the questionnaire's three income brackets. We use respondents aged 21–50, the sample age groups closest to the ages covered by the available population statistics; the population weights are

[3] For the United States, population estimates by single year of age allow us to match the sample age groups exactly in the first calculation; in the second, the top cell covers ages 41–49 against the sample age group 41–50. For Singapore and South Korea, official tabulations use ten- or five-year age groups, so we map ages 20–29, 30–39, and 40–49 to the sample age groups 21–30, 31–40, and 41–50, and ages 50 and over (Singapore) or 50–79 (South Korea) to the over-50 group.

the corresponding cell shares among childless adults under 50.

Online Appendix D.2 Population cell counts

United States. For each age–sex group, we estimate the number of childless adults by multiplying the U.S. Census Bureau's 2023 population count by the corresponding childlessness rate from the 2015–2019 National Survey of Family Growth (NSFG). These rates measure the share of women or men in each age group who have never had a biological child. The NSFG childlessness rates cover ages 15–49. For ages 50–79, which the survey does not cover, we assume childlessness rates of 15 percent for women and 19 percent for men, close to the shares observed at ages 40–49.[4] For the education and income cells, we use NSFG counts and childlessness rates by educational attainment and by household income relative to the federal poverty level. As an approximation, we map income bands below 150 percent, 150–299 percent, and 300 percent or more of the poverty level to the questionnaire's low, medium, and high income brackets, respectively.

Singapore. Official statistics report childlessness only for ever-married women. We therefore use marital status to approximate the population of childless adults. Specifically, we combine the Department of Statistics' 2024 tabulations of residents by marital status, age, and gender with data from the Census of Population 2020 on the number of children born to ever-married women by age group. We estimate the number of childless adults in each age group as the number of never-married residents plus the number of ever-married residents multiplied by the share of ever-married women in that age group who have had no children. We apply this share to both women and men. For the education cells, we use the same approach with the Census 2020 tabulations by highest qualification attained for ages 20–49, defining a degree as a university qualification.

South Korea. We use an analogous approach for South Korea. We combine the 2020 Population and Housing Census sample tabulation of Korean nationals by gender, five-year age group, marital status, and educational attainment[5] with age-specific shares of ever-married women who have had no children, reported in the census release. For the education cells, we use the same tabulation for ages 20–49, defining a degree as a completed four-year university degree or graduate study. Because childlessness among ever-married women is reported by age only, we apply the same age-specific

[4] Varying these assumed rates between 12 and 18 percent for women and between 15 and 23 percent for men yields U.S. *not-yet* share estimates ranging from 54.3 to 59.2 percent for ages 21–79.

[5] Korean Statistical Information Service (KOSIS), table `DT_1PM2003`.

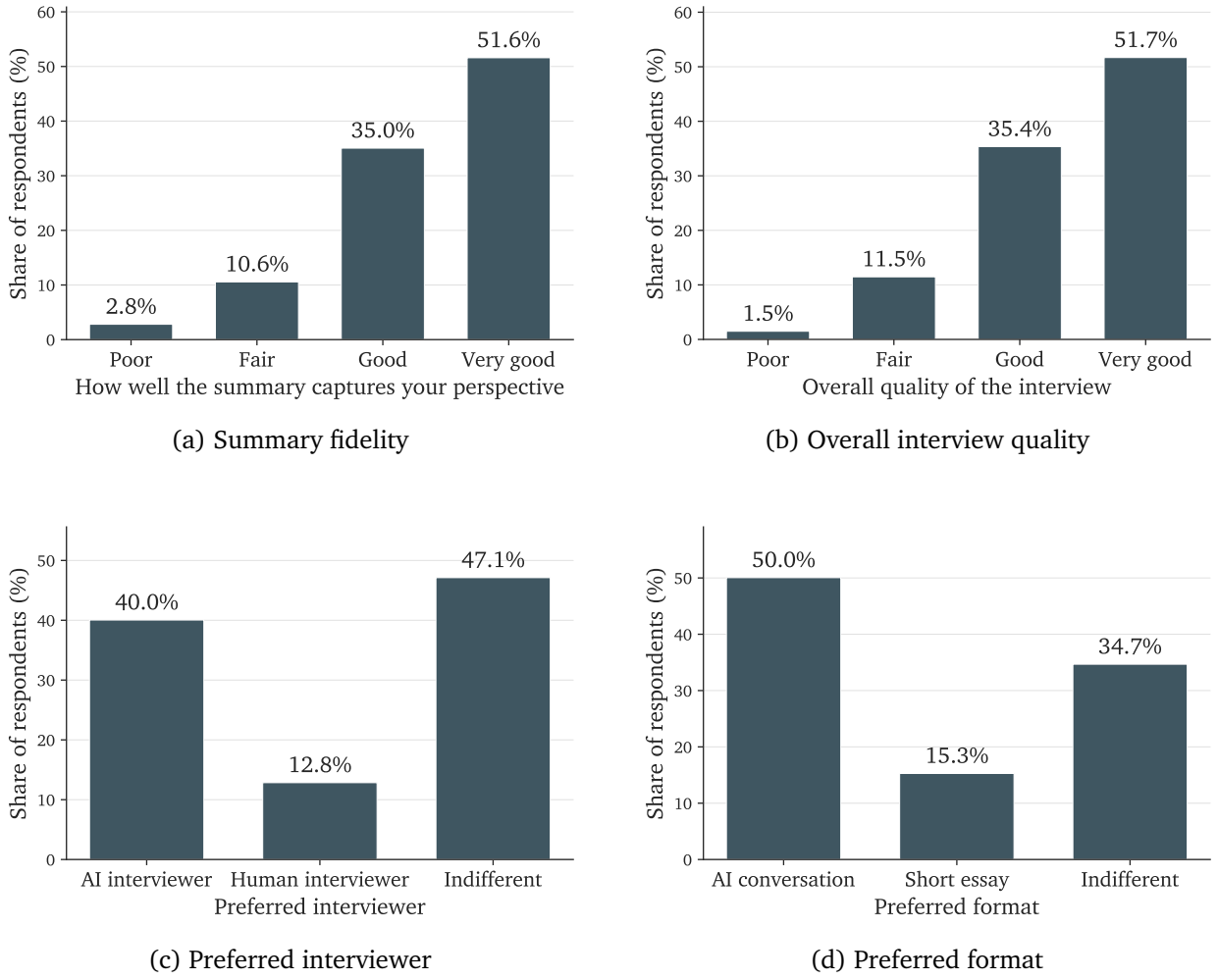
shares across education groups. No official tabulation by income is available for Singapore or South Korea.

Online Appendix D.3 Results and limitations

The reweighted estimates are close to the corresponding sample shares. In the full sample, 58.4 percent of respondents are classified as *not yet* (Table 1, "All respondents" row). After reweighting, the equally weighted average of the three country-specific estimates is 58.5 percent, with country-specific estimates ranging from 56.6 to 60.3 percent. For the younger target populations defined above, this average is 68.1 percent, with country-specific estimates ranging from 64.0 percent in South Korea to 72.5 percent in the United States (Table 5 Panel A). This increase reflects both the greater prevalence of *not-yet* respondents at younger ages and differences in age composition between our sample and the childless population. Reweighting by gender and education, or by gender and income, also changes the estimates little: among respondents aged 21–50, either approach changes the *not-yet* share by less than three percentage points (Table 5 Panel B).

These estimates are subject to three limitations. First, extrapolation assumes that, within each demographic cell, the sample share classified as *not yet* applies to the corresponding childless population. Reweighting cannot address selection within cells, including differences between online-panel participants and non-participants. Second, because our U.S. sample is restricted to respondents of East Asian ethnicity, the U.S. estimates combine cell-specific *not-yet* shares among East Asian Americans with the population composition of all childless U.S. adults. Third, our population counts for Singapore and South Korea use marital status to approximate childlessness. This approach treats never-married parents—a small group in both countries—as childless and assigns ever-married men the childlessness rate of ever-married women of the same age.

Subject to these limitations, the reweighting results provide little evidence that our sampling design mechanically inflates the prevalence of open childbearing decisions.



(a) Summary fidelity

(b) Overall interview quality

(c) Preferred interviewer

(d) Preferred format

Figure 1: Respondents' evaluation of the AI-led interview

This figure reports respondents' evaluations of the AI-led interview, based on MCQ2–MCQ5, the four hard-coded multiple-choice questions asked during the closing summary-and-evaluation stage. Panel (a) reports the results for MCQ2, the rating of the AI agent's summary fidelity on a four-point scale; Panel (b) reports the results for MCQ3, the rating of the overall interview quality on a four-point scale; Panel (c) reports the results for MCQ4, the preferred interviewer: AI, human, or indifferent; and Panel (d) reports the results for MCQ5, the preferred format: AI-led conversation, short essay, or indifferent. The exact wording of each question is reported in Online Appendix B.1.2.

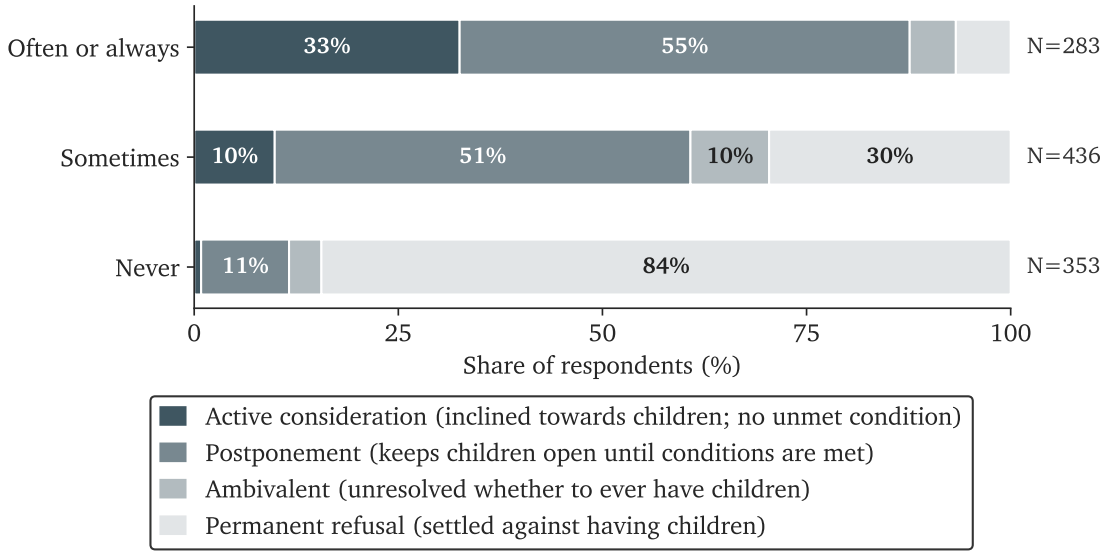


Figure 2: Transcript-based deliberation state by deliberation frequency

This figure reports, within each deliberation-frequency group, the distribution of the transcript-based deliberation-state classification. A large language model (GPT-5.2) reads each of the 1,074 transcripts—truncated immediately before the closing deliberation-frequency question, so the classifier never observes the respondent's answer—and classifies the deliberation state the respondent's reasoning reaches as active consideration, postponement, ambivalent, or permanent refusal (Section 2 of the main draft). Rows are defined by the answer to the closing question "How often do you find yourself thinking about having children?": never (N = 353), sometimes (N = 436), and often or always (N = 283); two respondents without a usable answer are excluded.

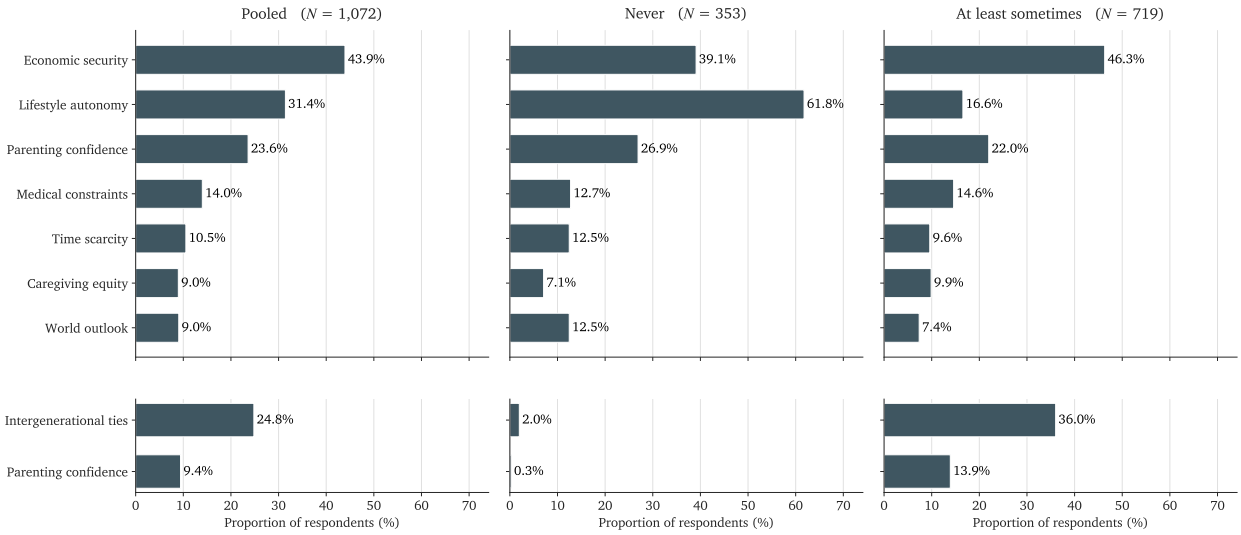


Figure 3: Narrative prevalence by self-reported deliberation frequency

This figure replicates Figure 2 of the main paper, replacing the transcript-based classification with the self-reported deliberation frequency. Groups are defined by the answer to the closing question "How often do you find yourself thinking about having children?": never ($N = 353$) versus at least sometimes, meaning sometimes, often, or always ($N = 719$). Two respondents without a usable answer are excluded; the pooled column covers the remaining 1,072. A narrative counts as prevalent for a respondent when it is "highly salient," meaning a salience score of 4 in the direction in which it pushes the respondent. All panels share the same scale.

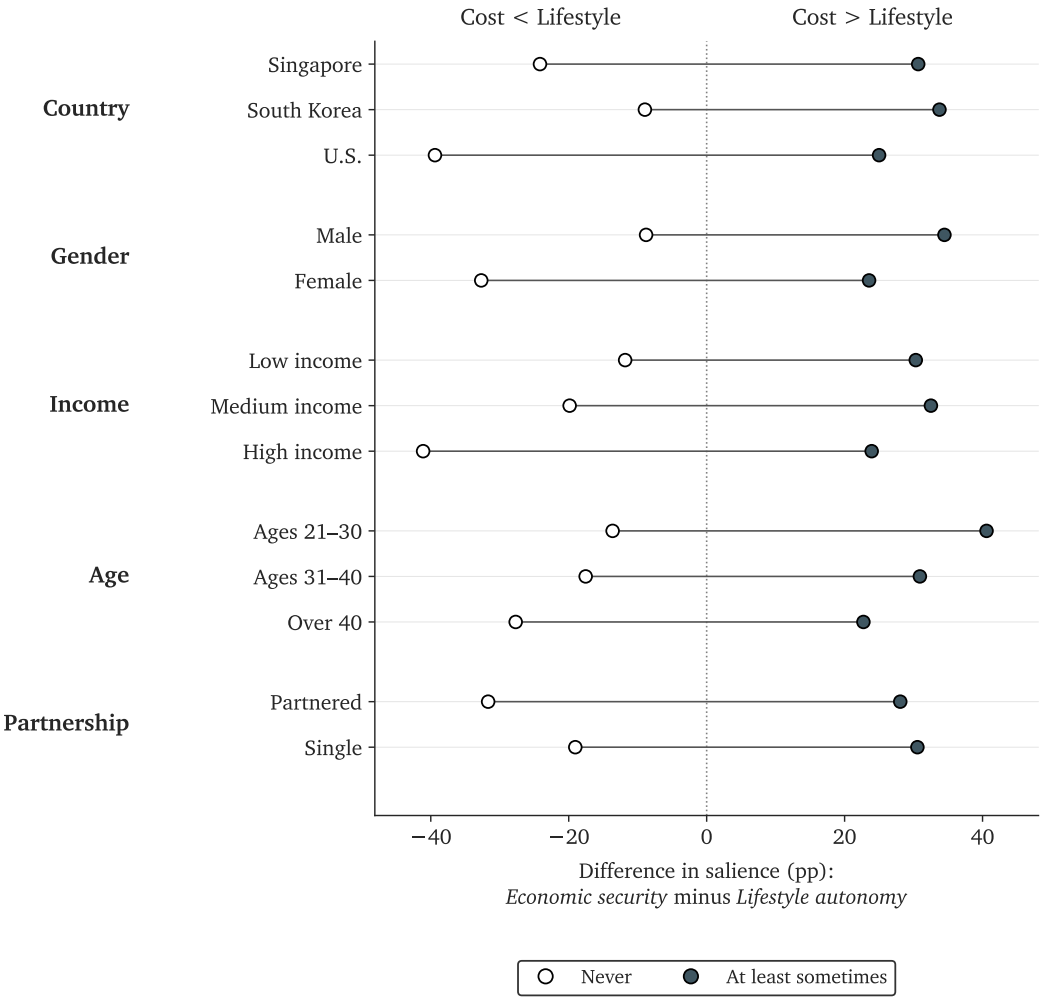
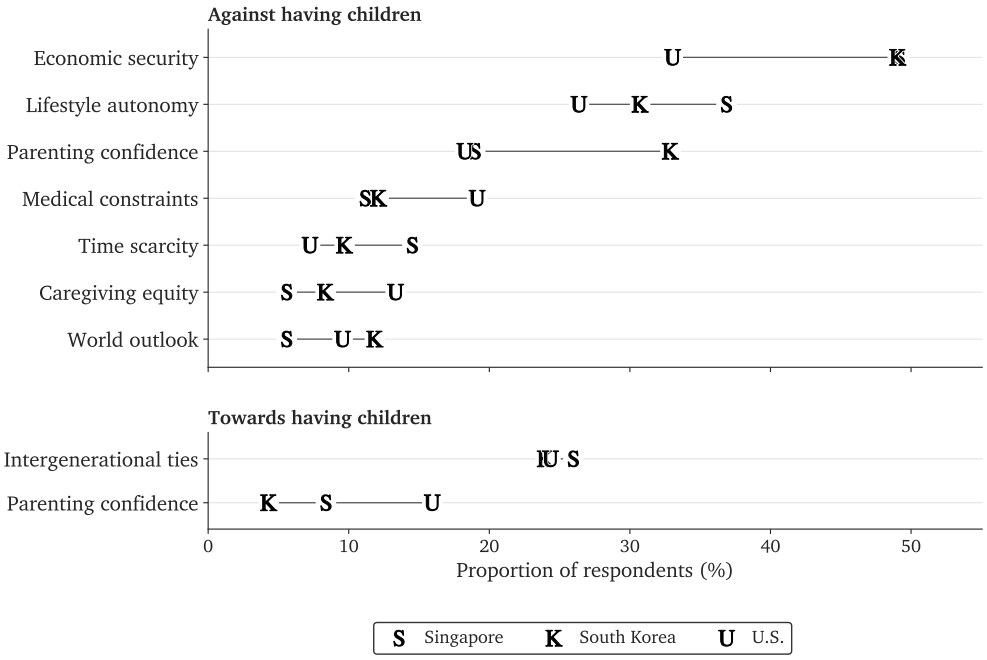
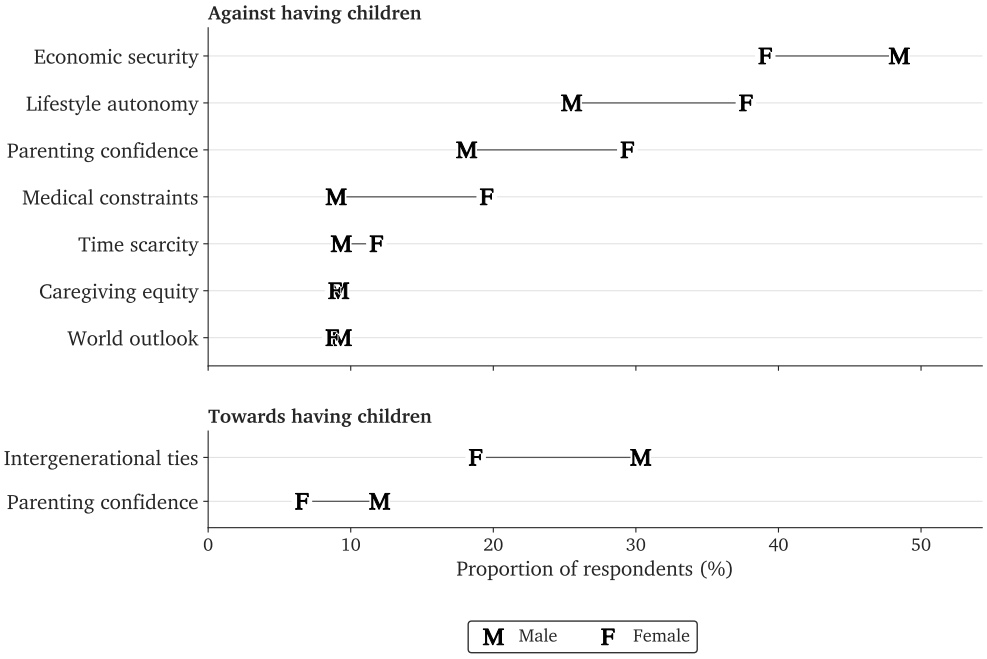


Figure 4: Cost-versus-lifestyle considerations under self-reported deliberation frequency

This figure replicates Figure 3 of the main paper, replacing the transcript-based classification with the self-reported deliberation frequency, grouped as in Figure 3. For each of 13 demographic subgroups, it plots the difference, in percentage points, between the share of respondents for whom *Economic security* is "highly salient" and the share for whom *Lifestyle autonomy* is "highly salient," separately for the self-reported *never* group (open circle) and the at-least-sometimes group (filled circle), with a line connecting the two estimates. Negative values indicate *Lifestyle autonomy* is more prevalent; positive values indicate *Economic security* is more prevalent. Subgroup definitions follow Figure 3 of the main paper. The inversion appears in all 13 subgroups.

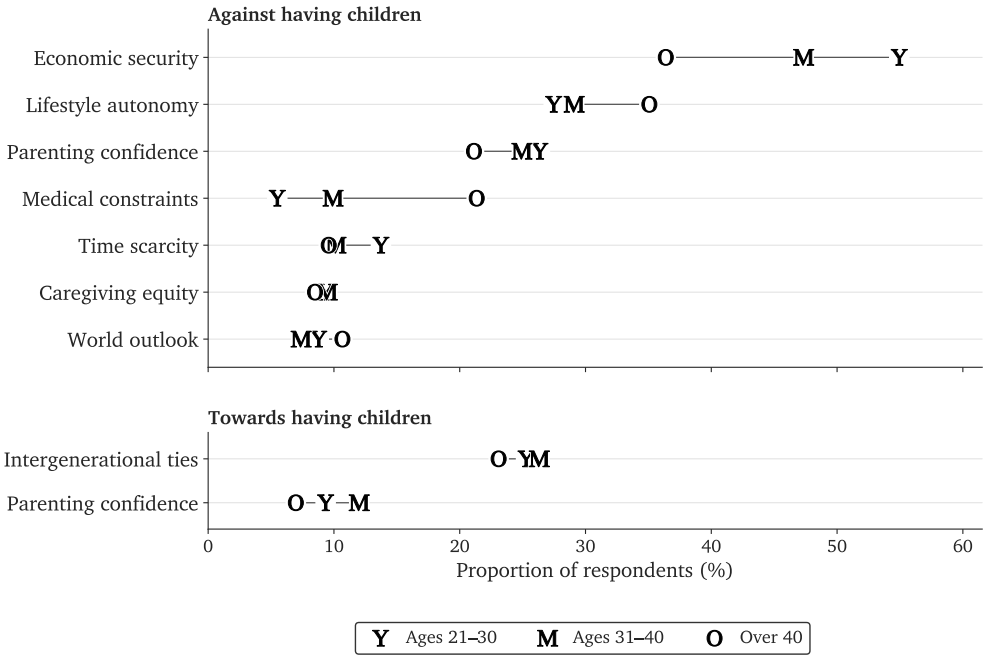


(a) By country

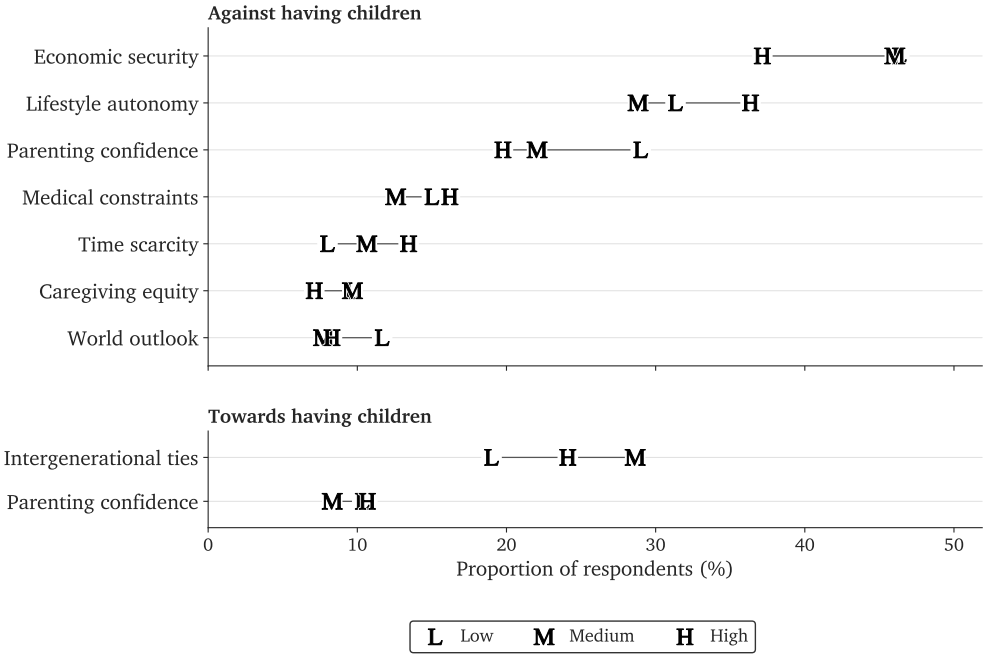


(b) By gender

Figure 5: Narrative prevalence by demographic group

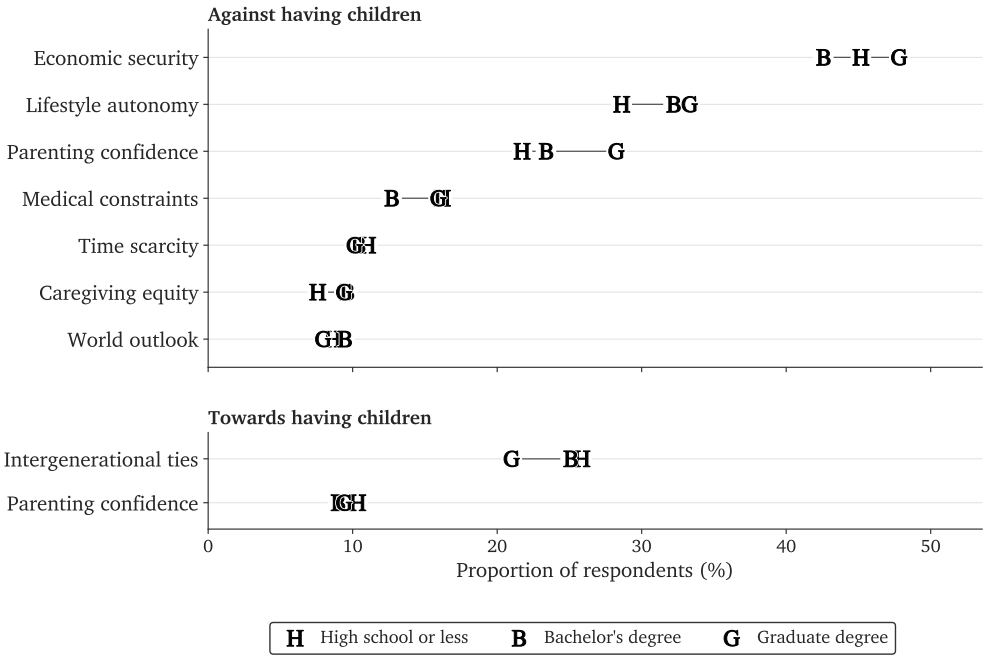


(c) By age

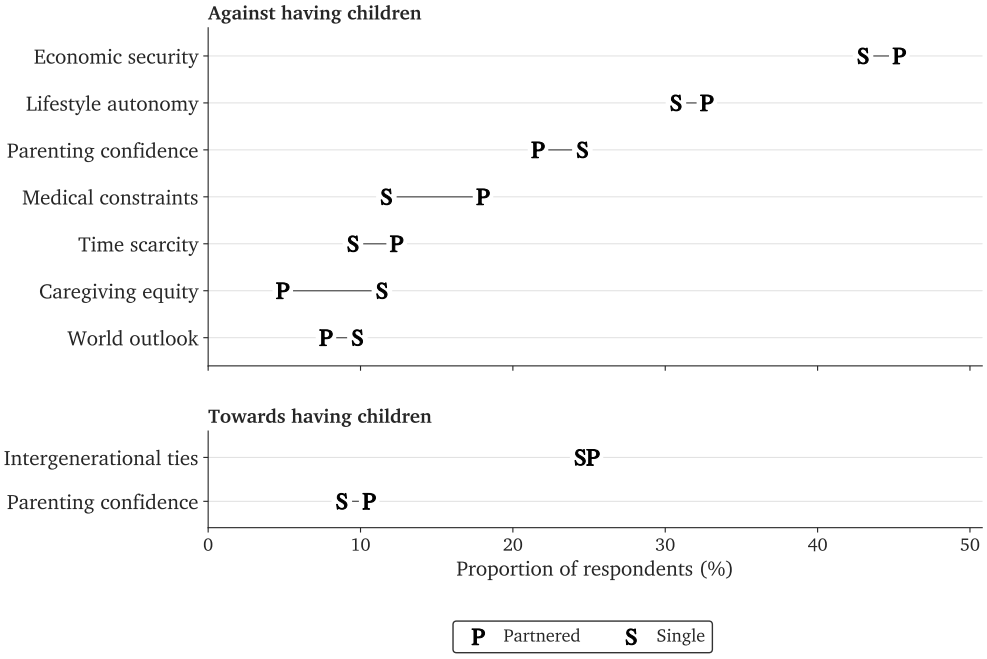


(d) By income

Figure 5: Narrative prevalence by demographic group

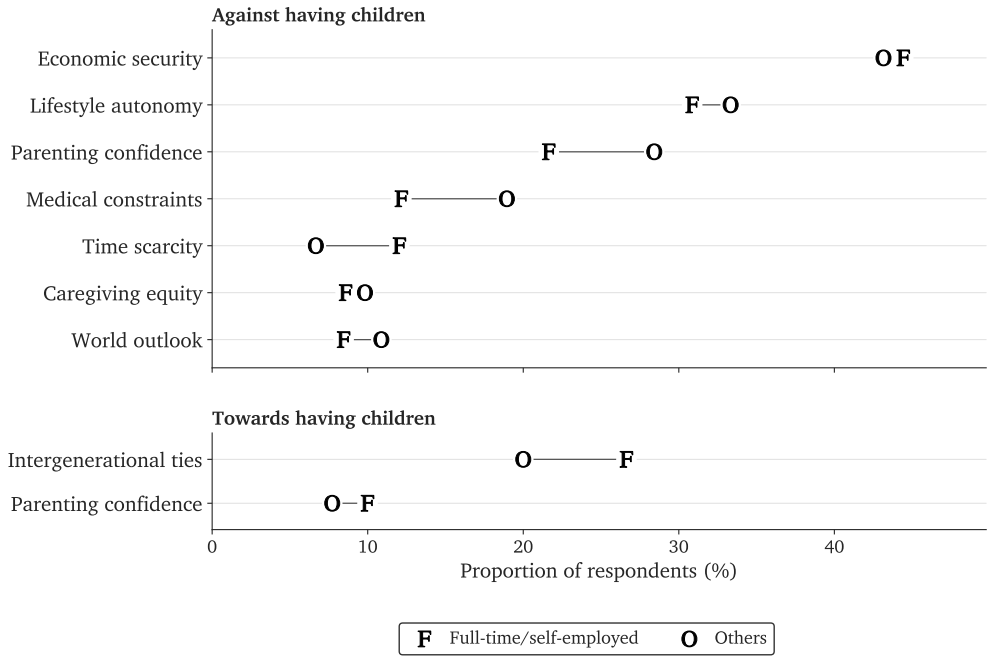


(e) By education



(f) By partnership status

Figure 5: Narrative prevalence by demographic group



(g) By employment status

Figure 5: Narrative prevalence by demographic group

This figure reports, for each of the nine key narratives, the share of respondents within each demographic group for whom the narrative is "highly salient," meaning a salience score of 4 in the direction in which it pushes the respondent, as in Figure 2 of the main paper. The upper panel of each plot reports the seven narratives pushing against having children and the lower panel the two pushing towards it, ordered by pooled prevalence; a horizontal line spans the range across groups. A letter marks each group, as keyed in the legend below each plot. Panels (a) through (g) split the 1,074 respondents by country, gender, age, income, education, partnership status, and employment status, respectively. Income brackets follow the country-specific cutoffs described in Table 1; "Partnered" denotes respondents who are married or currently in a relationship; "Full-time/self-employed" groups respondents employed full-time or self-employed, and "Others" all remaining employment statuses.

Table 1: Sample composition by deliberation state

The two columns under "Composition" report the demographic composition of the full sample of 1,074 childless respondents, showing the number and percentage of respondents in each demographic category. The two columns under "Deliberation state" report, within each category, the share of respondents in the *never* and *not-yet* groups, defined by the transcript-based deliberation state (Section 2 of the main draft). Income categories follow the country-specific brackets of the background information questionnaire (Online Appendix A): low is below approximately USD 2,300 per month, medium is between approximately USD 2,300 and USD 4,600, and high is above approximately USD 4,600, with parallel cutoffs in Singapore dollars and Korean won. "Partnered" denotes respondents who are married or currently in a relationship; "Single" denotes all other respondents. For *Income* and *Partnership*, we exclude four respondents who did not report their income and two who did not report their marital status, respectively.

	Composition		Deliberation state (row %)	
	N	%	Never	Not yet
Gender				
Male	557	51.9	33.4	66.6
Female	517	48.1	50.5	49.5
Age				
21 – 30	182	16.9	27.5	72.5
31 – 40	433	40.3	33.7	66.3
Over 40	459	42.7	54.7	45.3
Income				
Low	300	27.9	46.7	53.3
Medium	517	48.1	36.8	63.2
High	253	23.6	45.1	54.9
Education				
High school or less	290	27.0	43.8	56.2
Bachelor's degree	646	60.1	39.0	61.0
Graduate degree	138	12.8	49.3	50.7
Country				
Singapore	358	33.3	39.4	60.6
South Korea	371	34.5	42.6	57.4
U.S.	345	32.1	42.9	57.1
Partnership				
Partnered	388	36.1	40.2	59.8
Single	684	63.7	42.4	57.6
All respondents	1,074	100.0	41.6	58.4

Table 2: Salience of the two leading narratives by demographic subgroup and deliberation state

This table reports the percentage of respondents for whom *Lifestyle autonomy* and *Economic security* are "highly salient," separately for the *never* and *not-yet* groups, within each of the 13 demographic subgroups used in Figure 3 of the main paper. Figure 3 of the main paper plots the difference between these two percentages for each subgroup and group; this table reports the underlying levels. Subgroup definitions and the salience measure follow Section 2 of the main draft. "Partnered" denotes respondents who are married or currently in a relationship; "Single" denotes all other respondents. Two respondents who declined to report their marital status and four who declined to report their income are omitted from the corresponding subgroups. *N* is the number of respondents in each subgroup and group.

	Never			*Not yet*		
	N	Lifestyle autonomy	Economic security	*N*	Lifestyle autonomy	Economic security
Country						
Singapore	141	72.3	40.4	217	13.8	54.8
South Korea	158	57.6	44.3	213	10.8	52.6
U.S.	148	51.4	20.3	197	7.6	42.6
Gender						
Male	186	57.0	40.9	371	9.7	52.3
Female	261	62.5	31.0	256	12.5	47.3
Income						
Low income	140	53.6	39.3	160	11.9	52.5
Medium income	190	60.5	37.4	327	10.4	51.1
High income	114	67.5	26.3	139	10.8	46.0
Age						
Ages 21–30	50	64.0	42.0	132	13.6	59.8
Ages 31–40	146	65.8	44.5	287	10.5	48.8
Over 40	251	56.2	28.3	208	9.6	46.2
Partnership						
Partnered	156	63.5	35.9	232	12.1	51.7
Single	290	58.6	34.5	394	10.2	49.2
All respondents	447	60.2	35.1	627	10.8	50.2

Table 3: Argument-map structure by deliberation state

This table reports respondent-level structural statistics computed from the standardized argument maps, by the transcript-based deliberation state (Section 2 of the main draft): *never* (N = 447) and *not yet* (N = 627). Panel A classifies each respondent's map by which outcome nodes its directed paths reach: only *NotHavingChildren* (refusal only), only *HavingChildren* (having only), or both outcomes. Panel B restricts attention to respondents whose map contains at least one economic-domain consideration (domain tags are assigned when each DAG is generated; see Online Appendix C.1.1) and classifies which outcome those considerations reach. Shares in Panel B may not sum to 100 because a small number of economic considerations reach neither outcome (at most 0.4 percent). Reaching an outcome records the direction of the respondent's reasoning, not favorable or unfavorable sentiment. Argument-map construction is described in Section 2 of the main draft and detailed in Online Appendix C.1.

Panel A: Where the argument map leads (% of group)

	Refusal only	Having only	Both outcomes	*N*
never	60.4	0.0	39.6	447
not yet	2.7	30.1	67.1	627

Panel B: Where economic considerations lead (% of respondents raising any)

	Towards refusal	Towards having	Both outcomes	*N*
never	94.5	1.0	4.2	310
not yet	33.7	30.9	35.0	460

Table 4: Linear probability models of narrative salience on deliberation state

This table reports linear probability models of the highly salient indicators for *Economic security*, *Lifestyle autonomy*, and *Intergenerational ties* on an indicator for the transcript-based *not-yet* deliberation state (base group: *never*; Section 2 of the main draft). These estimates underlie the magnitudes reported in the results section of the main draft. Columns (2), (4), and (6) add the full set of demographic controls and report their coefficients: country (base: United States), gender (base: male), age band (base: ages 21–30), income bracket (base: low), education (base: high school or less), and partnership status (base: single). Columns (2), (4), and (6) exclude the six respondents who declined to report their income or marital status. Heteroskedasticity-robust standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

	Economic security		Lifestyle autonomy		Intergenerational ties	
	(1)	(2)	(3)	(4)	(5)	(6)
not yet	0.151***	0.113***	-0.493***	-0.502***	0.390***	0.392***
	(0.030)	(0.032)	(0.026)	(0.027)	(0.021)	(0.022)
Singapore		0.147***		0.115***		0.016
		(0.036)		(0.030)		(0.030)
South Korea		0.156***		0.046		0.001
		(0.037)		(0.030)		(0.030)
Female		-0.081***		0.035		-0.044*
		(0.030)		(0.025)		(0.024)
Age 31–40		-0.075*		-0.015		0.032
		(0.044)		(0.034)		(0.036)
Age over 40		-0.160***		-0.046		0.083**
		(0.043)		(0.035)		(0.036)
Medium income		-0.008		0.016		0.069**
		(0.037)		(0.030)		(0.028)
High income		-0.075*		0.052		0.055
		(0.045)		(0.036)		(0.034)
Bachelor's		-0.045		0.035		-0.039
		(0.036)		(0.031)		(0.029)
Graduate		0.062		-0.009		-0.051
		(0.053)		(0.042)		(0.040)
Partnered		0.027		0.015		-0.002
		(0.031)		(0.025)		(0.025)
Demographic controls	No	Yes	No	Yes	No	Yes
Adjusted R^2	0.022	0.062	0.274	0.286	0.197	0.202
N	1,074	1,068	1,074	1,068	1,074	1,068

Table 5: Share classified as *not yet*: Sample and population-reweighted estimates

Panel A reports the percentage of respondents classified as *not yet* under the transcript-based classification (Section 2 of the main draft), both in the sample and after reweighting by gender and age to the population of childless adults in each country. We report reweighted estimates for all adults aged 21–79 and separately for adults aged 21–49. "Three countries, equal weight" is the simple average of the three country-specific estimates. Panel B repeats the exercise for respondents aged 21–50, reweighting by gender and education (a bachelor's degree or higher versus below) and, for the United States, by gender and income bracket, to the corresponding population of childless adults under 50; dashes mark combinations for which no official population statistics are available. The four respondents who declined to report their income are omitted from the income cells of Panel D and from the income reweighting. Panels C and D report the cell-level inputs behind Panels A and B: the number of sample respondents, the percentage classified as *not yet*, and the cell's share of the country's childless adult population; Panel C reports this share for both target populations, all ages and under 50. Population weights are constructed from official population and childlessness statistics as described in Online Appendix D. Percentages are reported to one decimal place.

Panel A: Share classified as not yet after reweighting by gender and age (percent)

	Sample	Reweighted, ages 21–79	Reweighted, ages 21–49
United States	57.1	56.6	72.5
Singapore	60.6	58.7	67.7
South Korea	57.4	60.3	64.0
Three countries, equal weight	58.4	58.5	68.1

Panel B: Reweighting by gender and education or income, respondents aged 21–50 (percent)

	Sample, ages 21–50	Reweighted, gender × degree	Reweighted, gender × income
United States	63.5	65.0	63.8
Singapore	65.0	64.1	–
South Korea	60.8	62.7	–

Panel C: Cell-level inputs for Panel A (gender × age, all respondents)

		United States				Singapore				South Korea			
		N	*not yet*	Share (all)	Share (<50)	*N*	*not yet*	Share (all)	Share (<50)	*N*	*not yet*	Share (all)	Share (<50)
Female	21–30	28	64.3	19.8	26.8	38	73.7	17.3	22.9	23	60.9	21.8	25.0
Female	31–40	59	69.5	7.2	9.7	73	56.2	12.4	16.3	80	55.0	11.1	12.7
Female	41–50	60	28.3	4.5	5.6	41	43.9	7.5	9.9	51	47.1	5.4	6.2
Female	Over 50	19	0.0	11.5	–	23	21.7	13.4	–	22	27.3	5.4	–
Male	21–30	26	88.5	24.5	33.2	43	74.4	19.5	25.8	24	70.8	25.0	28.7
Male	31–40	70	70.0	12.1	16.4	72	79.2	12.5	16.4	79	69.6	15.3	17.5
Male	41–50	61	73.8	6.7	8.4	44	59.1	6.6	8.7	67	64.2	8.7	10.0
Male	Over 50	22	18.2	13.5	–	24	41.7	10.9	–	25	40.0	7.2	–

Panel D: Cell-level inputs for Panel B (gender × degree or income, respondents aged 21–50)

		United States			Singapore			South Korea		
		N	*not yet*	Pop. share	*N*	*not yet*	Pop. share	*N*	*not yet*	Pop. share
Female	No degree	44	61.4	21.0	27	59.3	18.6	24	70.8	25.6
Female	Degree	103	47.6	21.5	125	56.8	30.9	130	50.0	18.3
Female	Low income	31	64.5	10.8	37	59.5	–	48	52.1	–
Female	Medium income	76	48.7	11.0	73	63.0	–	83	57.8	–
Female	High income	39	48.7	20.8	40	45.0	–	22	40.9	–
Male	No degree	48	70.8	35.0	36	66.7	25.6	49	59.2	37.4
Male	Degree	109	76.1	22.6	123	74.0	24.9	121	71.1	18.7
Male	Low income	30	60.0	13.2	40	67.5	–	55	65.5	–
Male	Medium income	72	81.9	13.6	71	78.9	–	93	69.9	–
Male	High income	55	72.7	30.6	48	66.7	–	22	63.6	–

Online Appendix References

Balbo, N. and Barban, N. (2014). Does fertility behavior spread among friends? *American Sociological Review*, 79(3):412–431.

Becker, G. S. (1960). An economic analysis of fertility. In *Demographic and Economic Change in Developed Countries*, NBER Conference Series, pages 209–240. Columbia University Press, New York.

Chopra, F. and Haaland, I. (2023). Conducting qualitative interviews with AI. Working Paper 10666, CESifo.

Geiecke, F. and Jaravel, X. (2024). Conversations at scale: Robust AI-led interviews with a simple open-source platform. *The Review of Economic Studies*. Forthcoming.

Hwang, B.-H., Noh, D., and Shin, S. S. (2025). How investors pick stocks: Global evidence from 1,540 AI-driven field interviews. Working Paper 5844543, SSRN.

Kim, S., Tertilt, M., and Yum, M. (2024). Status externalities in education and low birth rates in Korea. *American Economic Review*, 114(6):1576–1611.

Kreyenfeld, M. (2010). Uncertainties in female employment careers and the postponement of parenthood in Germany. *European Sociological Review*, 26(3):351–366.

Lesthaeghe, R. (2010). The unfolding story of the second demographic transition. *Population and Development Review*, 36(2):211–251.

Miller, W. R. and Rollnick, S. (2012). *Motivational interviewing: Helping people change*. Guilford press.

Myong, S., Park, J., and Yi, J. (2021). Social norms and fertility. *Journal of the European Economic Association*, 19(5):2429–2466.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Thornton, A. (2005). *Reading History Sideways: The Fallacy and Enduring Impact of the Developmental Paradigm on Family Life*. University of Chicago Press, Chicago.