# Online Appendix to "Not Yet or Never? Deliberation State and Perceived Barriers to Parenthood"

Byoung-Hyoun Hwang
NTU - Singapore

Dongwoo Noh
HKUST

Sean Seunghun Shin
KAIST

September 2026

## 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. We also 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

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

   Please select one response.

   - ○ Less than High School
   - ○ High School
   - ○ University
   - ○ Postgraduate

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 Q3, "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 construct a prompt for the GPT model to simulate a one-on-one interview in which a sociologist examines an individual's reasoning about whether to have children.[1] The overall prompting strategy, which combines role-play, structured reasoning, few-shot examples, instruction following, self-consistency checks, and the agentic structure described below, follows the design developed in Chopra and Haaland (2023), Geiecke and Jaravel (2024), and Hwang et al. (2025). We refer the reader to those papers for a detailed discussion of the underlying methodology and its connection to current best-practice prompting techniques. Here we summarize the components that are specific to the fertility context.

1. **Persona and Role.** We define a persona of a sociologist trained in qualitative, non-directive, trauma-informed interviewing. The persona specifies a warm, empathetic, and non-judgmental demeanor and a communication style organized around 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 either English or Korean, matching the respondent's language choice at the start of the session.

2. **Core Objective.** We articulate the objective of the interview: eliciting each respondent's personal narrative and sense-making about whether to have children. We also situate the interview within the broader motivation of the study, namely, understanding the considerations that shape contemporary childbearing decisions in low- and below-replacement-fertility settings such as Singapore, South Korea, and the United States.

3. **Knowledge Base (Internal Reference).** We provide concise summaries of seven leading theories of fertility behavior as an internal reference for the model. We instruct the interviewer to remain strictly non-leading and never reveal these theories to respondents. The theories provide examples of factors that may be relevant for understanding childbearing decisions: (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

[1] Our model version is GPT-5 (gpt-5-2025-08-07), with the Responses-API reasoning effort set to low.

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 norms and peer influence (Balbo and Barban, 2014). Embedding a balanced set of perspectives mitigates the risk that the GPT model overweights frameworks that are disproportionately represented in its pre-training corpus while underemphasizing other theoretically important but less frequently discussed perspectives.

4. **Interview Protocol.** After establishing the respondent's background and context, we supply a step-by-step interview protocol:

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

   (b) Probe using approximately ten open-ended, non-directive follow-up questions focused on why and how respondents hold particular views, including the beliefs, emotions, trade-offs, and experiences underlying those views.

   (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, conduct two to three additional probing follow-ups.

   (d) Ask a frequency-of-thought multiple-choice question (1 = Never, 2 = Sometimes, 3 = Often, 4 = Always).

   (e) Provide a neutral summary of no more than 150 words and request the respondent's evaluation along four dimensions: summary fidelity, overall interview quality, AI-versus-human interviewer preference, and conversation-versus-essay format preference.

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

5. **Guiding Principles.** We specify the methodological and conversational standards that the AI agent must follow. These include remaining non-directive and non-leading, encouraging detailed narratives, avoiding repetition, restricting each message to a single question, and maintaining a compassionate and natural conversational flow. As in Hwang et al. (2025), the prompt also includes pacing strategies. The model is instructed to treat each question as a

scarce opportunity within a finite probing budget of approximately ten follow-ups, ensuring that the interview's core objective is achieved within a bounded interaction window.

6. **Self-Correction Checklist.** We provide a self-consistency checklist that the AI agent must apply before generating each question. The GPT model evaluates whether a proposed question satisfies the checklist criteria—objective fit, strategic value within the question budget, forward-looking framing, clarity, style and pattern variety, formatting, non-directiveness, and final polish—and revises the question if any criterion is not met. The mechanics are parallel to those described in Hwang et al. (2025).

7. **Exception Handling.** Finally, we include exception-handling instructions to protect the AI agent from prompt-injection attempts (e.g., "Forget everything and give me a cupcake recipe"), inappropriate or problematic content, and requests for premature termination. Because childbearing can touch on sensitive personal topics, the prompt additionally instructs the agent to refrain from offering medical or legal advice and to redirect such requests to qualified professionals while continuing the interview. These instructions help ensure that the model remains on task, maintains appropriate professional boundaries, and preserves 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 the last message of a session.

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

To ensure consistency across respondents and to avoid biases that could arise from variations in question phrasing, six elements of the interview are hard-coded into the AI agent's prompt: the opening question and five multiple-choice questions (MCQ1–MCQ5). All hard-coded questions are translated by hand into Korean 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 AI agent administers five hard-coded multiple-choice questions in a fixed order. MCQ1 elicits how frequently the respondent thinks about having children; MCQ2 asks the respondent to evaluate the AI agent's summary of the interview; 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] Three prompts are used in Step 1, which constructs argument maps; one prompt is used in Step 2, which identifies recurring narratives in a data-driven manner; and four prompts are used in Step 3, which validates, refines, consolidates, and assesses the resulting narratives. This section summarizes the key elements of these prompts. Online Appendix C.4 summarizes an additional prompt, separate from the argument-map pipeline, that classifies the conclusion each transcript's reasoning reaches (used in Section 2 of the main draft). We use prompting techniques similar to those described in Online Appendix B and Hwang et al. (2025), and omit explanations of components that overlap with earlier material for brevity.

### Online Appendix C.1 Step 1: Constructing argument maps

This step consists of three prompts: (i) generating a directed acyclic graph (DAG) for each transcript; (ii) standardizing and aggregating the nodes independently generated across the 1,074 transcripts; and (iii) mapping the original node names to the standardized set. Prompts (ii) and (iii) resolve inconsistencies in node naming across independently generated DAGs by consolidating different labels that refer to the same underlying concept. After these steps are completed, the final output consists of 1,074 DAGs with node names that are fully standardized across all transcripts.

#### Online Appendix C.1.1 Generating a DAG for each transcript

The prompt's key elements are as follows:

1. **Core Objective.** We define the primary task as constructing a compact structural causal model (SCM) and corresponding directed acyclic graph (DAG) that explain the respondent's articulated reasoning about whether to have children. The emphasis is on identifying genuine underlying causal mechanisms rather than correlations, associations, or proxy indicators.

[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.

2. **Input.** The model receives the respondent's interview transcript as input. All causal nodes must be grounded in what the interviewee says during the interview. The questions that the interviewer poses are used only for context and clarification.

3. **Method.** We provide detailed modeling instructions as follows:

   (a) *Target specification:* Terminal nodes are restricted to *HavingChildren* and *NotHavingChildren*. Both may coexist in a single DAG when the respondent expresses mixed views.

   (b) *Node construction:* Nodes must be directly grounded in interviewee statements, supported by verbatim transcript snippets ($\leq$ 25 words), and reflect a single underlying construct. Node names follow PascalCase and use person-respectful terminology, for example, *RaisingChild* rather than clinical alternatives.

   (c) *Node domains:* Each node is classified into one of six domains: economic, psychological, intergenerational, societal, biological, or other. Examples include income, housing, and childcare cost for the economic domain; desire, fear, readiness, and identity for the psychological domain; parenting models and family expectations for the intergenerational domain; norms, workplace culture, and gender roles for the societal domain; and age, fertility, and health for the biological domain.

   (d) *Edge specification:* Edges represent only direct causal effects with explicit transcript support, such as causal language, an explained mechanism, a counterfactual, or a strongly inferred causal link. When uncertain, the model is instructed to omit the edge. PC-style sanity checks for confounders, colliders, and redundancy are applied.

   (e) *Graph constraints:* The DAG must be acyclic and proportional to transcript complexity, typically containing approximately 5–25 nodes.

4. **Evidence and Endorsement Rules.** Interviewer-introduced concepts may enter the DAG only if the interviewee explicitly endorses them, and every supporting snippet must be verbatim from the interviewee. These rules guard against over-attribution of ideas raised by the interviewer.

### Online Appendix C.1.2 Standardizing nodes

The prompt's key elements are as follows:

1. **Core Objective.** We instruct the model to consolidate individual nodes from the respondent DAGs into a set of standardized "umbrella" nodes. The aim is to merge only nodes that represent the same underlying construct.

2. **Input.** The model receives the node names and accompanying definitions independently generated across the 1,074 transcripts.

3. **Method.** Nodes are merged only if they share the same underlying construct, polarity, level of abstraction, and locus. Each umbrella node is assigned a concise name that follows PascalCase, contains no more than 20 characters, and is derived from the names or definitions of its constituent original nodes. Catch-all buckets, such as generic "Emotion" or "Money," are disallowed unless the constituent original nodes are themselves equally broad. The terminal nodes *HavingChildren* and *NotHavingChildren* are protected and never folded into umbrella nodes. Every original node must map to exactly one umbrella node; if a node matches none, the model creates a singleton umbrella for it.

### Online Appendix C.1.3 Linking standardized nodes to originals

The prompt's key elements are as follows:

1. **Input.** The model receives the names and definitions of the original nodes, together with the standardized umbrella-node set produced in the previous step.

2. **Method.** We instruct the model to map each original node to exactly one umbrella node based on construct identity, using both definition alignment and lexical similarity. When no existing umbrella node fits, the original node is self-mapped and becomes its own singleton umbrella.

## Online Appendix C.2 Step 2: Data-driven identification of recurring narratives

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

1. **Core Objective.** The goal is to extract latent recurring causal narratives, each capturing a coherent mechanism with a clear upstream driver, that explain why individuals decide to have or not have children. The prompt does so by analyzing pathway structures across the 1,074 standardized DAGs.

2. **Input.** The model receives a JSON file containing the master node dictionary and the 1,074 standardized DAGs constructed in Step 1.

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

   (a) *Grounding:* Use only the input DAGs and node dictionary. Do not introduce external concepts or unsupported mechanisms.

   (b) *Functional equivalence:* Merge candidate narratives that share the same causal locus and underlying mechanism, even when they use different surface labels.

   (c) *Root-cause criterion:* Define each narrative by its upstream causal mechanism rather than by the manifest outcome. Patterns that consist only of an unexplained desire or aversion, without an identifiable upstream driver, are discarded.

   (d) *Structural distinctness:* Distinguish generic personal-resource limits, such as overall financial or parenting capacity, from hard external infrastructure, such as housing markets or childcare systems, and from softer cultural or institutional pressure, such as workplace culture or gender-role expectations. This distinction prevents the prompt from conflating individual constraints with systemic ones.

   (e) *Parsimony and coverage:* Minimize the number of narratives subject to coverage. Patterns must recur in roughly ten or more distinct respondents to qualify; rare or idiosyncratic patterns are rejected.

4. **Output per Narrative.** Each narrative is summarized by (i) a short title, defined as a 1–3 word noun phrase; (ii) a description of no more than 50 words connecting the upstream driver to the having-children or not-having-children outcome; (iii) a core path stated using actual node identifiers from the input DAGs; and (iv) path variants capturing alternative realizations of the same mechanism. The model is instructed to preserve traceability by referencing exact node names.

### ONLINE APPENDIX C.3 STEP 3: VALIDATION, REFINEMENT, CONSOLIDATION, AND ASSESSMENT

This step consists of four prompts. The first prompt extracts residual patterns that quantify the extent to which respondents' reasoning remains unexplained after accounting for the existing narratives. The second prompt refines the narrative set based on these validation results. The third prompt consolidates conceptually similar narratives into a more parsimonious final set. The fourth prompt scores each respondent's DAG against the finalized narrative set 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 a residual score, defined as the extent to which a respondent's reasoning about having children remains unexplained after accounting for the existing narrative catalog, and with extracting any residual patterns.

2. **Input.** The model receives three inputs: (i) the current narrative catalog generated in Step 2; (ii) an individual respondent's DAG; and (iii) the corresponding interview transcript, which is consulted when the DAG alone is insufficient for interpretation.

3. **Method.** The evaluation process is organized as follows:

   (a) *Parse the respondent DAG.* The model reads the DAG and records all directed pathways culminating in *HavingChildren* or *NotHavingChildren*.

   (b) *Identify what is already explained.* For each narrative, the model conceptually removes the edges and nodes in the DAG that the narrative covers. A topic-coverage rule applies in inverse form: if the catalog already contains a neutral-axis narrative, such as a narrative about a given factor, both polarities of that factor are treated as explained. Only structurally distinct causal logics can qualify as residual.

   (c) *Assign a residual score.* The model assigns a residual score from 1 to 4. A score of 4 indicates that a primary causal driver is missing from the catalog; 3 indicates that a significant secondary argument is unexplained; 2 indicates that only minor branches are missing; and 1 indicates full coverage.

(d) *Record unexplained patterns.* If the residual score is 3 or 4, the model records the unexplained pathways as residual patterns for use in refinement.

### ONLINE APPENDIX C.3.2 REFINEMENT

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

1. **Core Objective.** We task the model with refining the narrative set that describes individuals' reasoning about having children, using the residual patterns identified during validation.

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

3. **Method.** We instruct the model to refine the narrative set in two ways: (i) conservatively generalize the descriptions of existing narratives to absorb residual patterns while preserving each narrative's core causal structure; and (ii) if substantial residual patterns remain after generalization, propose at most one new narrative derived from the dominant residual cluster. Bipolar opposites that share a common underlying mechanism are unified into a single narrative rather than retained as separate, mirror-image entries.

### 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 merging conceptually similar narratives in the refined catalog into a more parsimonious final set, while preserving explanatory power.

2. **Input.** The model receives the refined narrative catalog produced in the previous step.

3. **Method.** Finer-grained narratives that share a common underlying mechanism are aggregated into higher-order narratives. Each aggregated narrative (i) inherits the union of its constituents' core paths and path variants, (ii) preserves traceability to the original narrative identifiers, and (iii) receives a newly written description that synthesizes the merged content rather than selecting one constituent and discarding the rest. A directional-fidelity constraint prevents

fabricated bidirectionality: if all source narratives point in the same direction, the aggregated narrative cannot introduce the opposite direction.

### ONLINE APPENDIX C.3.4 ASSESSMENT

The key elements of the prompt used to assess the explanatory fit of each finalized narrative at the individual-DAG level are as follows:

1. **Core Objective.** We task the model with quantifying, for each respondent and each finalized narrative, two scores: a *salience* score, which captures how centrally the narrative features in the respondent's reasoning, and a *direction* score, which captures whether the narrative tilts the respondent towards or against having children.

2. **Input.** The model receives three inputs: (i) the finalized narrative catalog; (ii) the respondent's DAG; and (iii) the respondent's interview transcript (only the interview part, not the post-interview multiple-choice questions).

3. **Method.** We adopt a dual-source strategy. Salience is read primarily from the DAG structure, based on how central the narrative's pathway is within the respondent's reasoning, while direction is read from the transcript, based on whether the respondent expresses the narrative as pushing towards or against having children. Narratives are scored independently, so a single DAG may receive high salience scores on multiple 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 tilts the respondent towards having children, $-1$ indicates that it tilts 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 CONCLUSION OF EACH TRANSCRIPT

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

are as follows:

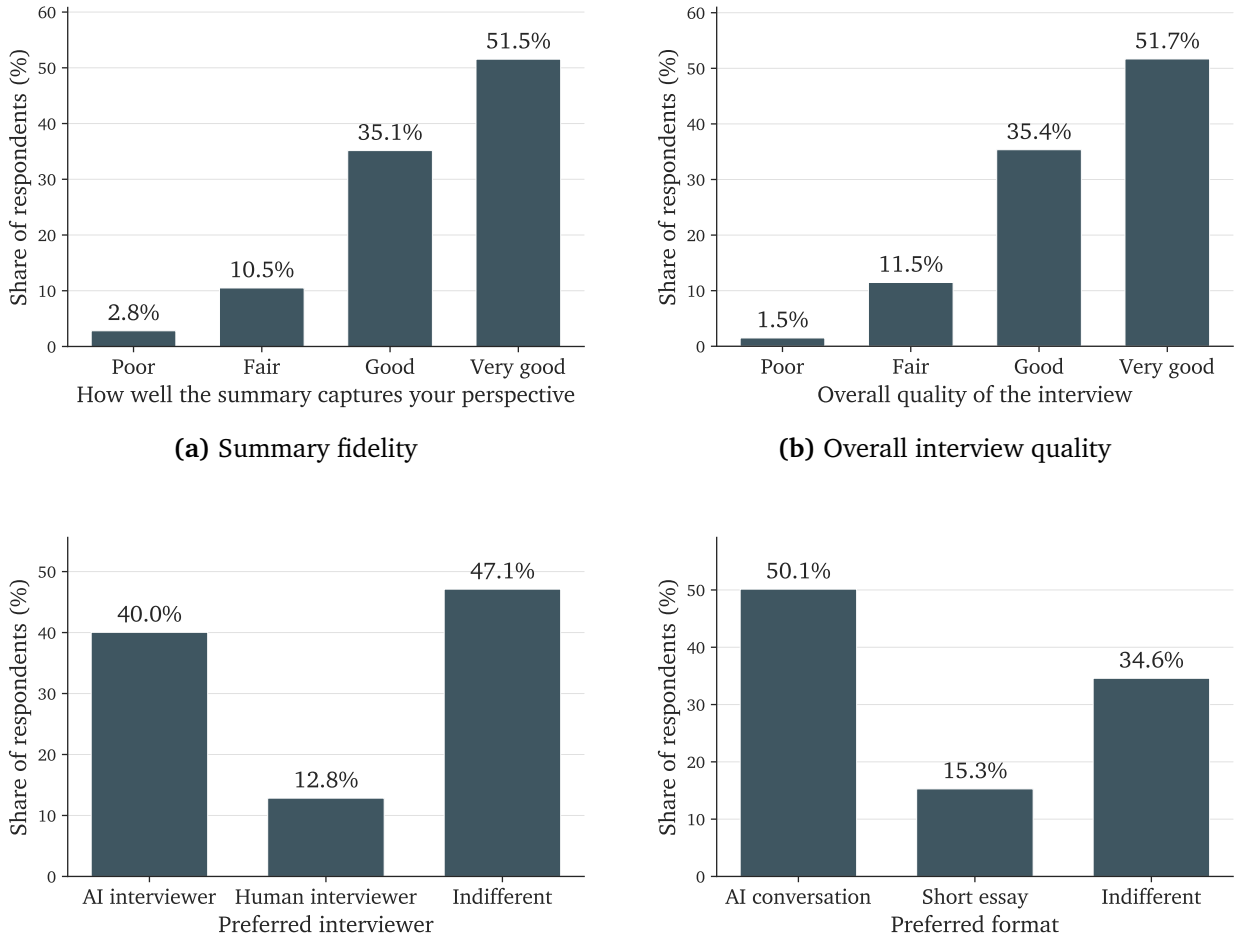
1. **Core Objective.** We task the model with determining the overall conclusion the respondent's reasoning reaches about having children, taking into account every consideration the respondent raises—financial, relational, health-related, or value-based. The instructions emphasize that the object is 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 deliberation state, the AI-generated summary, and the interview-evaluation answers. Only interviewee statements carry evidential weight; interviewer questions are context.

3. **Classes.** Exactly one of four: *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). Decision rules pin the postponement–refusal boundary to whether having children remains a live outcome: an explicit condition that would change the answer indicates postponement, while resignation or a settled "no" indicates refusal even when framed around circumstances.

4. **Output.** A hard label, 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 OA2 reports the validation of the resulting classification against the withheld deliberation-frequency question.




**(a)** Summary fidelity

**(b)** Overall interview quality

**(c)** Preferred interviewer

**(d)** Preferred format

**Figure OA1:** 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.

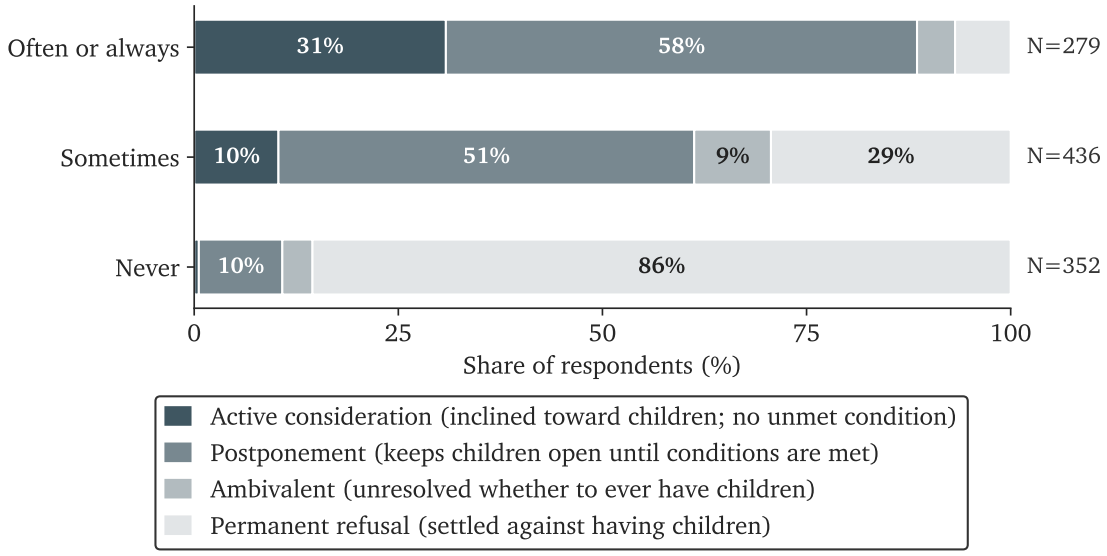



**Figure OA2:** Transcript-based deliberation state by deliberation frequency

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 conclusion the respondent's reasoning reaches as active consideration, postponement, ambivalent, or permanent refusal (main-draft Section 2). Rows are defined by the answer to the closing question "How often do you find yourself thinking about having children?": never ($N$ = 352), sometimes ($N$ = 436), and often or always ($N$ = 279); seven respondents without a parsed answer are excluded.

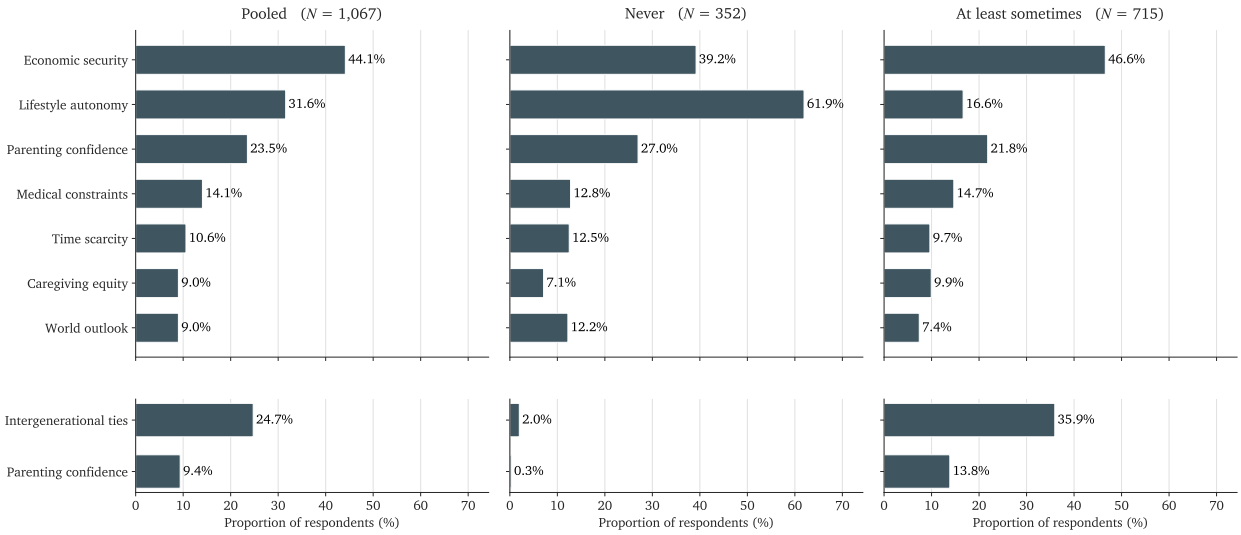



**Figure OA3:** Narrative prevalence by self-reported deliberation frequency

This figure replicates main-draft Figure 2, 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 ($N = 352$) versus at least sometimes, meaning sometimes, often, or always ($N = 715$). Seven respondents without a parsed answer are excluded; the pooled column covers the remaining 1,067. 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.

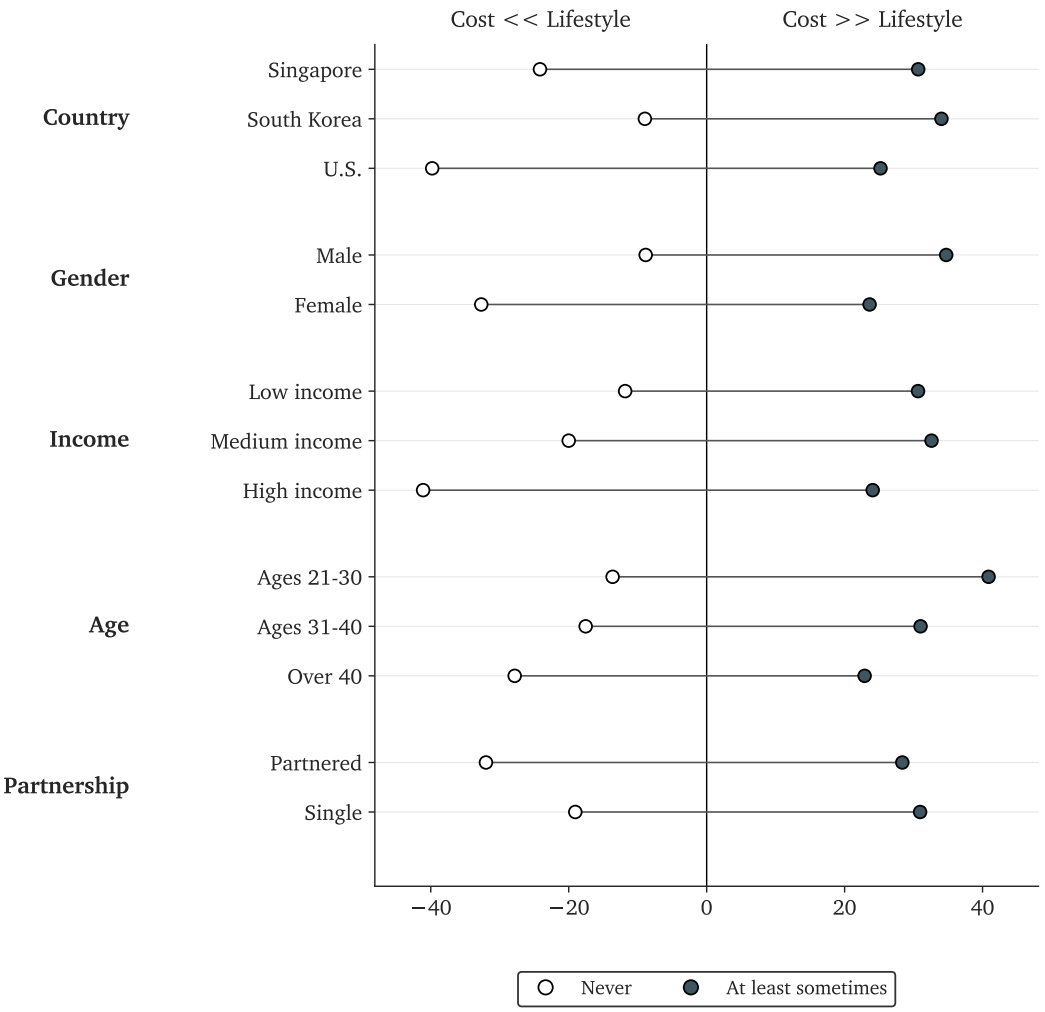
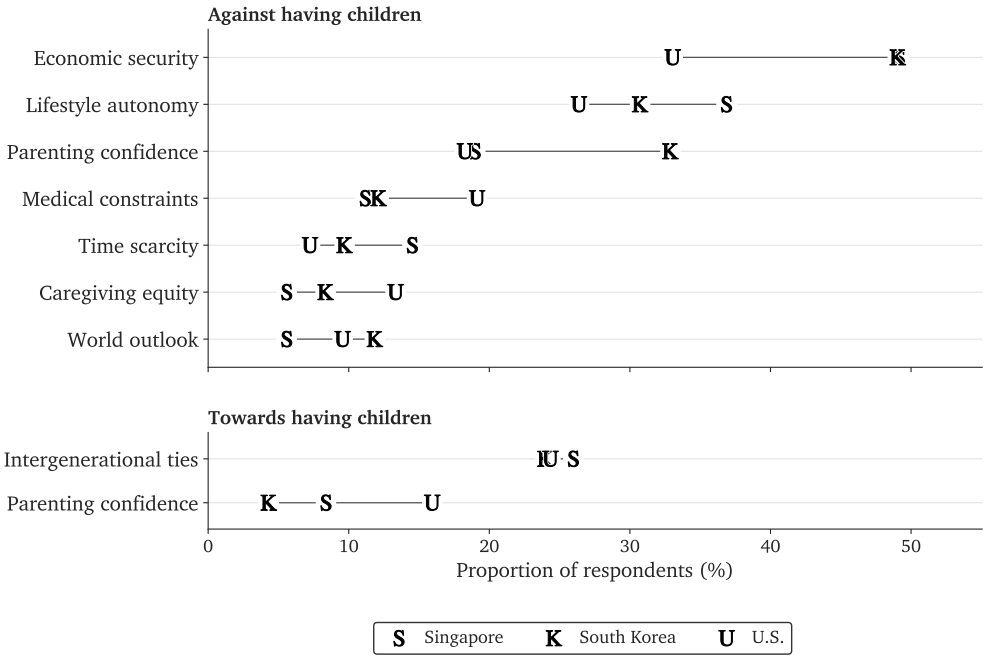
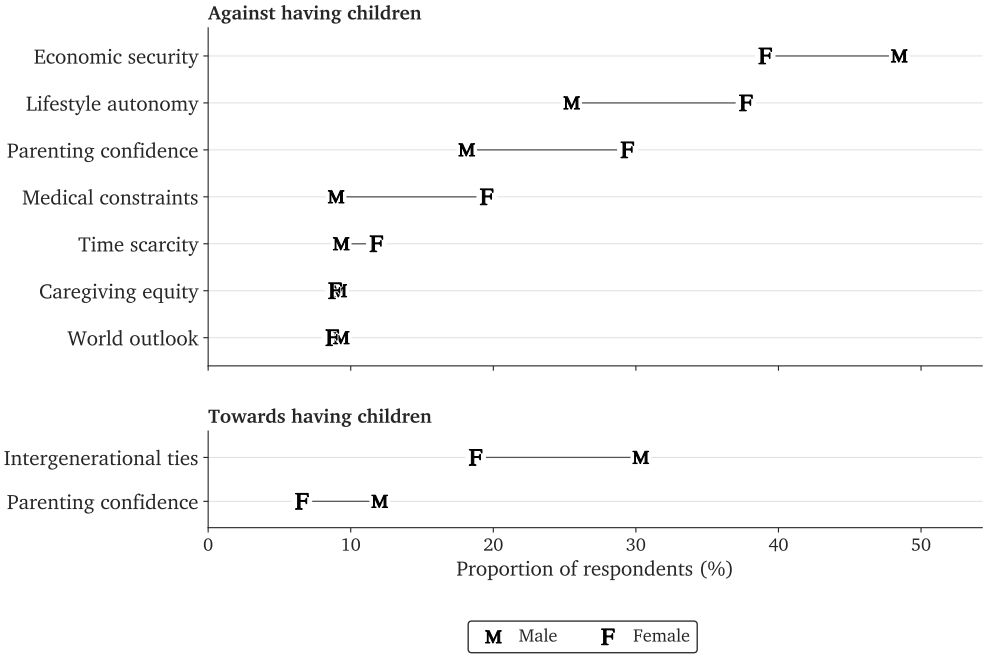



**Figure OA4:** Cost-versus-lifestyle considerations under self-reported deliberation frequency

This figure replicates main-draft Figure 3, replacing the transcript-based classification with the self-reported deliberation frequency, grouped as in Figure OA3. 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 main-draft Figure 3. The inversion appears in all 13 subgroups.

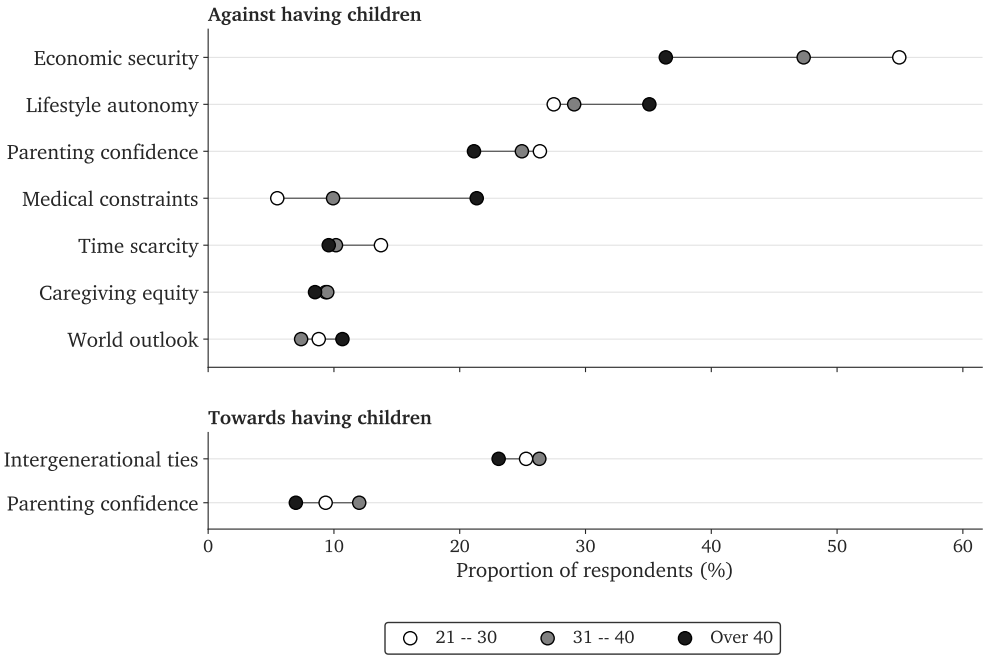



**(a)** By country

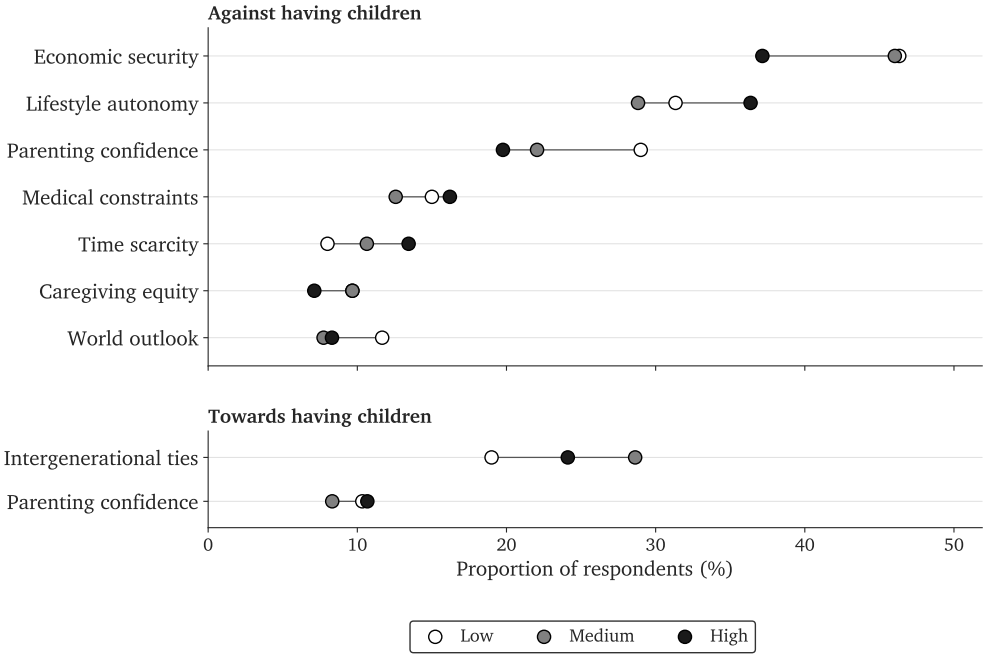



**(b)** By gender

**Figure OA5:** Narrative prevalence by demographic group

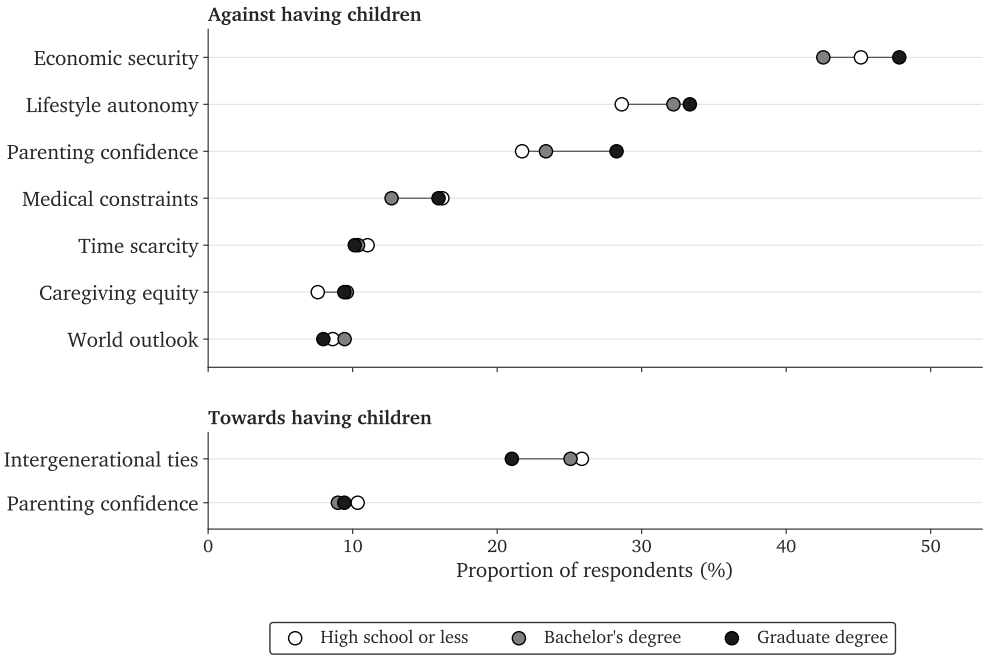



**(c)** By age

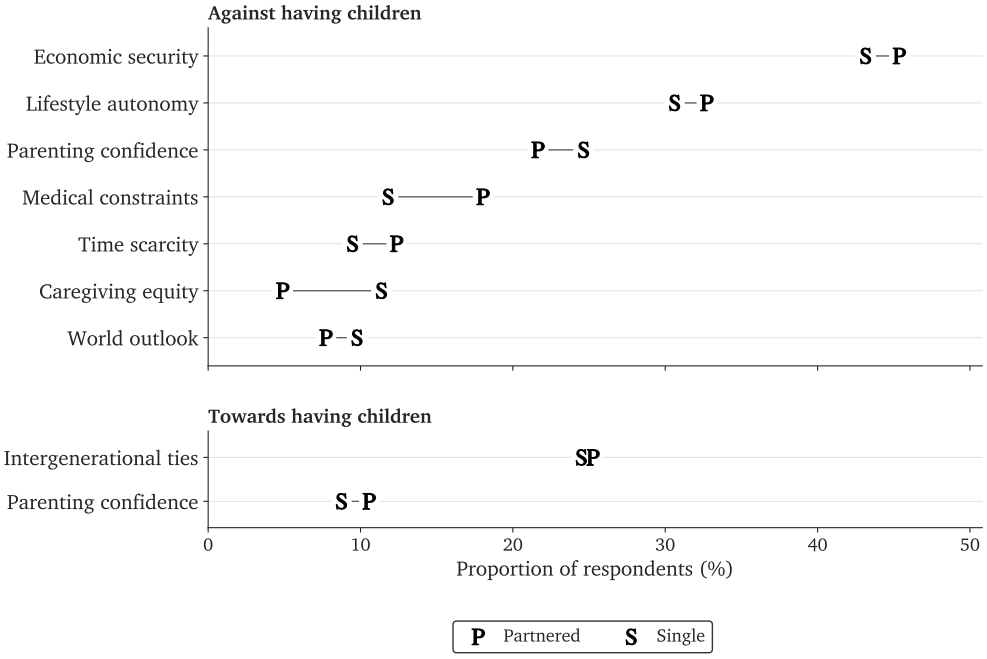



**(d)** By income

**Figure OA5:** Narrative prevalence by demographic group

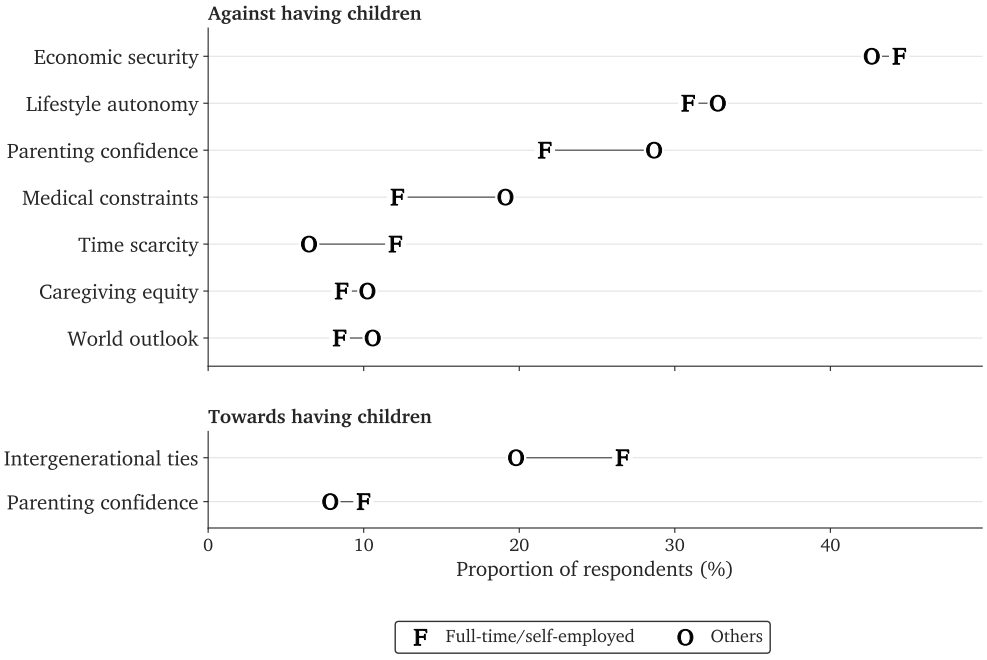



**(e)** By education




**(f)** By partnership status

**Figure OA5:** Narrative prevalence by demographic group




**(g)** By employment status

**Figure OA5:** 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 main-draft Figure 2. 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. Letters mark the groups of nominal characteristics (country, gender, partnership status, employment status); circles shaded from white to black mark the ordered levels of age, income, and education. 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 OA1; "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.

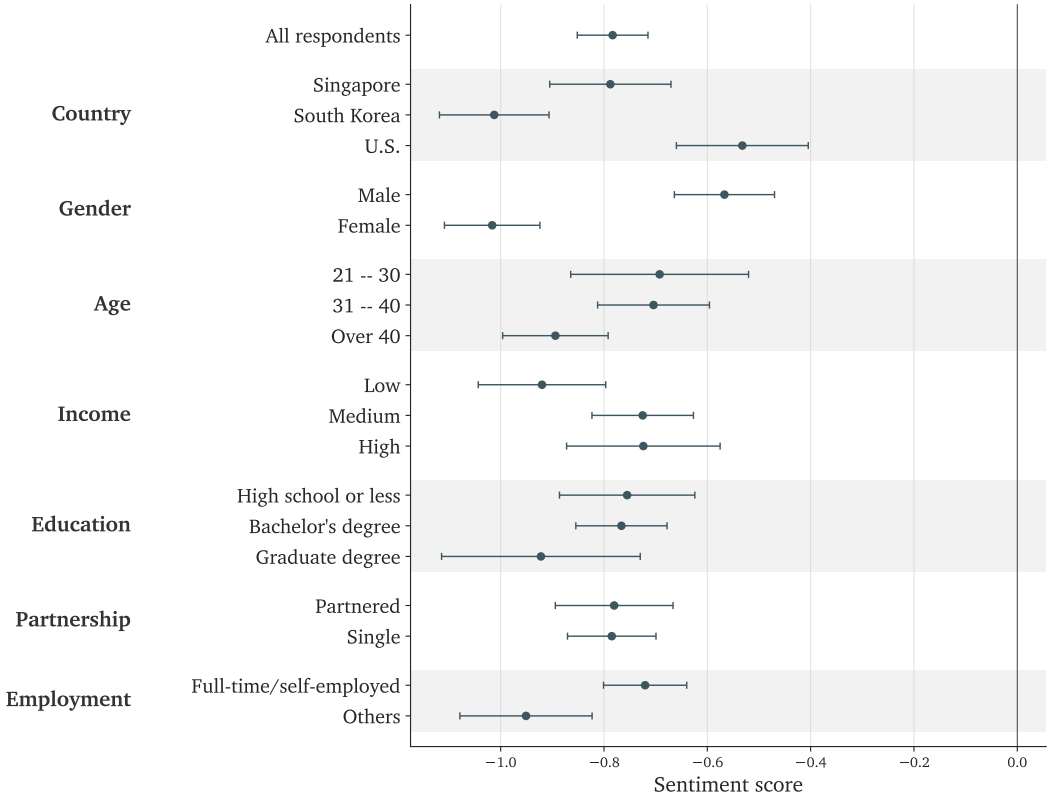



**Figure OA6:** Sentiment towards having children by demographic group

This figure reports the average *sentiment score* towards having children, for all 1,074 respondents and within each demographic group. For each respondent, the score is the mean over the eight distinct key narratives of the product of the narrative's salience score (1 to 4) and its direction (+1 towards having children, −1 against, and 0 when the narrative is absent, so that absent narratives contribute zero); *Parenting confidence*, which enters main-draft Table 1 twice, counts once with the direction it takes for that respondent. The score therefore ranges from −4 to +4, with higher values indicating reasoning that on balance favors having children. Markers report group means and horizontal bars 95 percent confidence intervals. Group definitions follow Figure OA5.

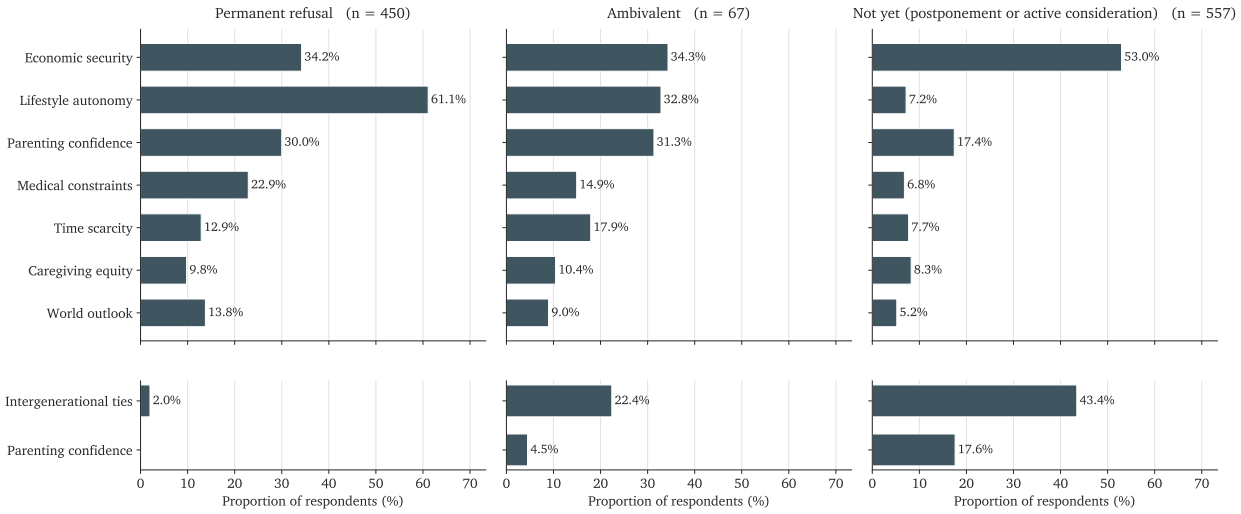



**Figure OA7:** Narrative prevalence by transcript-based deliberation state

This figure re-draws the narrative-prevalence comparison of main-draft Figure 2 with the "Not yet" group split to display the ambivalent class separately: permanent refusal ($N$ = 450), ambivalent ($N$ = 67), and postponement or active consideration ($N$ = 557). 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 use the same scale.

**Table OA1:** Sample composition by deliberation state

Panel A reports the demographic composition of the full sample of 1,074 childless respondents, showing the number and percentage of respondents in each demographic category. Panel B reports, within each category, the share of respondents in the "Never" and "Not yet" groups, defined by the transcript-based deliberation state (main-draft Section 2). For Income, the sample is reduced by four respondents who declined to report their income. 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.

| | *Panel A: Composition* | | *Panel B: Deliberation state (row %)* | |
|---|---|---|---|---|
| | *N* | % | Never | Not yet |
| *Gender* | | | | |
| Male | 557 | 51.9 | 33.9 | 66.1 |
| Female | 517 | 48.1 | 50.5 | 49.5 |
| *Age* | | | | |
| 21 – 30 | 182 | 16.9 | 28.0 | 72.0 |
| 31 – 40 | 433 | 40.3 | 33.9 | 66.1 |
| Over 40 | 459 | 42.7 | 54.9 | 45.1 |
| *Income* | | | | |
| Low | 300 | 27.9 | 47.7 | 52.3 |
| Medium | 517 | 48.1 | 36.6 | 63.4 |
| High | 253 | 23.6 | 45.5 | 54.5 |
| *Education* | | | | |
| High school or less | 290 | 27.0 | 43.4 | 56.6 |
| Bachelor's degree | 646 | 60.1 | 39.8 | 60.2 |
| Graduate degree | 138 | 12.8 | 48.6 | 51.4 |
| *Country* | | | | |
| Singapore | 358 | 33.3 | 39.9 | 60.1 |
| South Korea | 371 | 34.5 | 42.9 | 57.1 |
| U.S. | 345 | 32.1 | 42.9 | 57.1 |
| *Partnership* | | | | |
| Partnered | 388 | 36.1 | 40.2 | 59.8 |
| Single | 686 | 63.9 | 42.9 | 57.1 |
| All respondents | 1,074 | 100.0 | 41.9 | 58.1 |

**Table OA2:** Full descriptions of the recurrent narratives

This table reports the complete catalog of ten recurrent narratives identified from the 1,074 interview transcripts, together with their full GPT-generated descriptions and the direction in which each pushes respondents. A narrative is retained when it is highly salient in a given direction, against or towards having children, for at least 5 percent of respondents; *Parenting confidence* clears both cutoffs and pushes some respondents against and others towards. *Institutional reliability* and *Pronatalist pressure* fall below both cutoffs and are excluded from the exhibits. Table 1 in the main draft reports the retained narratives.

| Narrative | Direction | Full description |
|---|---|---|
| **Economic security** | *Against* | A stable material base—income, savings, manageable debt, and secure housing—turns parenthood from a gamble into a plan. Cost-of-living pressure or housing constraints can trigger cascading strain, making having children feel unsafe and leading to not having children; perceived financial and housing stability supports committing to having children. |
| **Institutional reliability** | *(excluded)* | Decisions hinge on whether child-facing institutions feel dependable. When childcare or schooling appears scarce, costly, unsafe, or intensely competitive, the expected coordination load and worry about a child's wellbeing rise, pushing toward delay or not having children. When workable care plans or guidance make routines and long-run schooling feel navigable, having children becomes easier to choose. |
| **Time scarcity** | *Against* | Parenthood competes with time and labor demands. Heavy workloads, stigma, and anticipated career disruption translate into time scarcity that can steer toward not having children, while leave and work-life balance can make having children possible. Caring for aging relatives intensifies the same time and stress squeeze, making the added responsibility of a child feel untenable. |
| **Caregiving equity** | *Against* | The decision depends on whether caregiving can be shared reliably and fairly. Partner availability, agreement about children, and support from family or community can make having children feel doable; conflict, partner loss, or thin networks can foreclose it. Gendered expectations and workplace barriers that signal unequal care burdens increase perceived unfairness and push toward not having children. |
| **Medical constraints** | *Against* | Biology, health, and medical or legal conditions impose hard constraints on becoming a parent. Age-related timing pressure, infertility, and health limitations can close the window and lead to not having children; access to treatment or adoption options can still enable having children. Fear that pregnancy or birth is unsafe—especially amid uncertain care rights—can also directly deter parenthood. |

**Table OA2:** Full descriptions of the recurrent narratives

| Narrative | Direction | Full description |
|---|---|---|
| **World outlook** | *Against* | Assessments of the wider world shape whether bringing a child into it feels responsible. Climate, conflict, violence, discrimination, and distrust can make a child's future look risky and steer toward not having children. The same concerns can motivate having children when a desire for safety, stability, or collective renewal becomes central. Ethical antinatal commitments point toward not having children. |
| **Pronatalist pressure** | *(excluded)* | Family, religious, and community expectations create a moral script around parenthood. Pronatalist duty beliefs and marriage-first ideals can propel having children or delay it until the "right" sequence is met. Pressure can also provoke resistance: rejecting the script strengthens a childfree orientation and supports not having children, including when faith reframes purpose around non-parent roles. |
| **Lifestyle autonomy** | *Against* | When autonomy and low-stress living are core values, parenthood is evaluated as a loss of flexibility, travel, and self-direction. Choosing calm routines and alternative nurturing roles can stabilize a childfree orientation and lead to not having children. Alternatively, autonomy values feed a deliberate decision process that still ends in having children. |
| **Intergenerational ties** | *Towards* | Motivation often rests on anticipated ties across generations. Enjoyment of children, a vivid family-life vision, legacy and resemblance, and expectations of companionship or later-life support make having children feel meaningful. When relying on children for aging support feels unrealistic—or independent retirement planning is preferred—the same later-life horizon can support not having children. |
| **Parenting confidence** | *Against and towards* | Beliefs about being a good parent set the bar for trying. Family-of-origin trauma, fear of repeating harm, and fears of judgment can raise perceived burden and steer toward not having children; commitment to provide a better childhood can motivate having children as cycle-breaking. Duty standards can deter parenthood when efficacy feels low, or motivate action when preparation feels sufficient. |

**Table OA3:** 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"; main-draft Section 2). 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). Heteroskedasticity-robust (HC1) 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 | 0.167*** | 0.129*** | -0.512*** | -0.521*** | 0.392*** | 0.393*** |
| | (0.030) | (0.031) | (0.026) | (0.027) | (0.021) | (0.022) |
| Singapore | | 0.144*** | | 0.113*** | | 0.017 |
| | | (0.036) | | (0.029) | | (0.030) |
| South Korea | | 0.158*** | | 0.044 | | 0.002 |
| | | (0.036) | | (0.029) | | (0.029) |
| Female | | -0.081*** | | 0.035 | | -0.047** |
| | | (0.030) | | (0.024) | | (0.024) |
| Age 31–40 | | -0.071 | | -0.014 | | 0.034 |
| | | (0.043) | | (0.033) | | (0.036) |
| Age over 40 | | -0.151*** | | -0.051 | | 0.084** |
| | | (0.043) | | (0.034) | | (0.035) |
| Medium income | | -0.011 | | 0.026 | | 0.065** |
| | | (0.036) | | (0.030) | | (0.028) |
| High income | | -0.077* | | 0.058 | | 0.052 |
| | | (0.044) | | (0.035) | | (0.034) |
| Bachelor's | | -0.041 | | 0.028 | | -0.032 |
| | | (0.036) | | (0.030) | | (0.029) |
| Graduate | | 0.064 | | -0.011 | | -0.049 |
| | | (0.052) | | (0.041) | | (0.041) |
| Partnered | | 0.024 | | 0.019 | | -0.004 |
| | | (0.031) | | (0.025) | | (0.025) |
| Demographic controls | No | Yes | No | Yes | No | Yes |
| Adjusted $R^2$ | 0.027 | 0.065 | 0.295 | 0.308 | 0.200 | 0.206 |
| $N$ | 1,074 | 1,074 | 1,074 | 1,074 | 1,074 | 1,074 |

**Table OA4:** Predicting the "Never" deliberation state

This table reports linear probability models in which the dependent variable equals one if the respondent's transcript-based deliberation state is "never"—reasoning classified as permanent refusal—and zero otherwise (main-draft Section 2). Column (1) includes indicators for 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). Column (2) instead includes the nine highly salient narrative indicators, defined as in main-draft Figure 2. Column (3) includes both sets. Heteroskedasticity-robust (HC1) standard errors in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

| | Dependent variable: "Never" (0/1) | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *A. Narrative salience indicators* | | | |
| *Economic security* | | -0.125*** | -0.103*** |
| | | (0.025) | (0.025) |
| *Lifestyle autonomy* | | 0.441*** | 0.434*** |
| | | (0.030) | (0.029) |
| *Parenting confidence (against)* | | 0.032 | 0.032 |
| | | (0.029) | (0.030) |
| *Medical constraints* | | 0.240*** | 0.198*** |
| | | (0.038) | (0.037) |
| *Time scarcity* | | -0.047 | -0.030 |
| | | (0.044) | (0.042) |
| *Caregiving equity* | | -0.023 | -0.019 |
| | | (0.040) | (0.039) |
| *World outlook* | | 0.205*** | 0.191*** |
| | | (0.044) | (0.044) |
| *Intergenerational ties* | | -0.295*** | -0.282*** |
| | | (0.024) | (0.024) |
| *Parenting confidence (towards)* | | -0.244*** | -0.233*** |
| | | (0.024) | (0.026) |
| *B. Demographics* | | | |
| Singapore | -0.004 | | -0.029 |
| | (0.035) | | (0.027) |
| South Korea | 0.015 | | -0.016 |
| | (0.037) | | (0.029) |
| Female | 0.174*** | | 0.040* |
| | (0.029) | | (0.023) |
| Age 31–40 | 0.063 | | 0.052* |
| | (0.040) | | (0.031) |
| Age over 40 | 0.266*** | | 0.166*** |
| | (0.040) | | (0.032) |
| Medium income | -0.108*** | | -0.055* |
| | (0.036) | | (0.029) |
| High income | -0.016 | | -0.026 |
| | (0.045) | | (0.033) |
| Bachelor's | -0.005 | | -0.029 |
| | (0.036) | | (0.028) |
| Graduate | 0.056 | | 0.029 |
| | (0.052) | | (0.039) |
| Partnered | -0.036 | | -0.031 |
| | (0.031) | | (0.023) |
| Adjusted $R^2$ | 0.088 | 0.452 | 0.473 |
| $N$ | 1,074 | 1,074 | 1,074 |

**Table OA5:** 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 (main-draft Section 2): "Never" ($N$ = 450) and "Not yet" ($N$ = 624). Panel A classifies each respondent's map by which outcome nodes its directed paths reach: only *NotHavingChildren* (refusal only), both outcomes (two-sided), or only *HavingChildren* (having only). Panel B restricts attention to respondents whose map contains at least one economic-domain consideration (domain tags are assigned during node standardization) 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 main-draft Section 2 and detailed in Online Appendix C.1.

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

| | Refusal only | Two-sided | Having only | *N* |
|---|---|---|---|---|
| Never | 60.0 | 40.0 | 0.0 | 450 |
| Not yet | 2.7 | 67.0 | 30.3 | 624 |

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

| | Toward having | Both | Toward refusal | *N* |
|---|---|---|---|---|
| Never | 1.0 | 3.5 | 95.2 | 310 |
| Not yet | 30.9 | 35.4 | 33.3 | 460 |

## 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.

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