Supplementary Online Material (SOM):

Generic References to Gender Predict Essentialism and Stereotyping Even When They Express Counter-Stereotypic Ideas

Josie Benítez[1], Emily Foster-Hanson[2], and Marjorie Rhodes[1]

[1] New York University
[2] Swarthmore College

Corresponding author: Josie Benítez

Email: josiebenitez@nyu.edu

ORCiD: https://orcid.org/0000-0003-3385-1509

Supplementary Online Materials (SOM)

Parent-child Conversation Task

Methods

In the main text, we include demographic information for the 192 families that completed the initial language session, summarize the study design, and briefly describe our coding protocol. Here, we provide further details on our transcription and speech coding procedures. Although transcription and speech coding was conducted on the full set of (parent and child) utterances, the analyses in this manuscript focus on *parent* language only.

Transcription procedure. All transcription and speech coding was conducted using the Datavyu video coding software (2014). The transcription strategy was adapted from the rigorous procedure used by the PLAY project (Adolph et al., 2019). The initial phase of transcription coding involved marking the timed onset and offset of each utterance, and identifying who the speaker was (child, parent, other). A valid utterance was defined as any unit of speech separated by grammatical closure, intonation, or prolonged pause (the statement "*Look at this girl! She's painting her nails*," for example, would be coded as two distinct utterances). Only communicative utterances made by real humans were transcribed; any voices originating from a person not present on camera, voice assistants, or instances where non-communicative sounds were made (e.g., unintelligible speech, whistling, sighs), were excluded from transcription.

Speech coding procedure. After transcription was complete, a team of trained research assistants conducted speech coding in a series of iterative passes. All coders were trained with the first author using sample videos; all valid session videos (*n* = 192,

100.00%) were subjected to review by a second coder who addressed any discrepancies in transcription and speech coding by discussing with the coding team. Given the unmoderated remote nature of our study administration, it was expected that conversations would deviate slightly. As such, all videos first underwent *focus coding*, which involved determining whether a given utterance was on-task, off-task, or an instance of a parent reading the prompting questions out-loud. Utterances deemed by coders as reasonably relating to the task were considered on-task (focused); utterances that sought clarification (e.g., "*Huh*?", "*Right*?") or any language not related to the storybook (e.g., "*Don't bother your sister*", "*Next page*") were coded as "off-task" (not focused); across conversations, 88.28% of all utterances were identified as being "on-task" (focused). On-task utterances were then subjected to person-reference coding to determine whether the utterance explicitly referred to a person (e.g., "Yes, he can do laundry"), or not (e.g., "What's the name of that tool?"; "That house is a nice color"); on-task utterances involving third-parties (e.g., "Batman," "Superheroes") were also marked as referring to a person. Overall, 59.87% of all on-task utterances were identified as referring to a person and were retained for speech coding.

On-task utterances that explicitly referred to a person were coded for how they referred to gender (generics, specific gender labels, pronouns) and the stereotypic content they communicated (stereotype-congruent, -incongruent, -neutral). Utterances were coded as generic if they referred to an abstract category of people (e.g., "Mommies cook dinner", "A fireman helps fight fires") or specific if they referred to an individual (e.g., "Jenny likes cooking", "My daddy is a fireman"). Utterances were coded as specific labels if they explicitly labeled the gender of the character in the storybook in

a specific manner (e.g., “That’s a daddy,” “What is that girl doing?”). Coding for gender stereotype-congruent and -incongruent language involved evaluating whether the content of a given utterance was congruent (e.g., “Only mommy puts on makeup”), incongruent (e.g., “That boy can paint his nails too, if he wants”; “Anyone can play with trucks if they want to”) or neutral (e.g., “Moms eat breakfast”; all other utterances) with regards to the expression of cultural gender stereotypes (for the full coding manual, please see our project page on the Open Science Forum: https://osf.io/htn7z/).

Results

In the manuscript, we describe how the linguistic features (generics, specific labels) and content (stereotype-congruent and incongruent talk) used in parents’ natural language varied as a function of the stimuli (gender and stereotype-consistency) depicted. Here, we describe (child participant) age-related changes and include additional results for comparative analyses concerning the stereotypic content of parents’ generics, and descriptive statistics on the use of hedging (e.g., “Girls *can* play football”, “Boys *sometimes* wear pink”) in parents’ generic language.

Parents’ Generic References to Gender. Parents of boys generated more generics as their children got older (β = 0.36, *95% CI* [0.07, 0.65]), while production of generics among parents of girls did not vary across child age in this sample (β = -0.07, *95% CI* [-0.34, 0.20]; interaction of child gender and mean-centered age, Wald $X^2(1)$ = 4.74, p = .03). Regardless of their child’s gender, parents generated fewer specific labels (β = -0.46, *95% CI* [-0.69, -0.26]; main effect of child age, Wald $X^2(1)$ = 19.47, p < .001) as their children got older.

Stereotypic Content of Parents’ Generics. Families used more

stereotype-neutral generics (e.g., “Moms eat breakfast”; *M* = 1.73, *95% CI* [1.57, 1.90]) than stereotype-congruent generics (e.g., “Boys are not supposed to dance”; *M* = 0.22, *95% CI* [0.15, 0.29]; contrast $p < .001$) or stereotype-incongruent generics (e.g., “Girls can dig for worms”; *M* = 0.29, *95% CI* [0.22, 0.36]; contrast $p < .001$; main effect of stereotypic content on generics, Wald $X^2(2) = 923.55$, $p < .001$), and families’ overall production of stereotype-congruent and -incongruent generics did not differ significantly (contrast $p = .18$). Families’ production of stereotype-congruent, -incongruent, and -neutral generics did not vary across child age in this sample ($p = .32$).

Use of Hedging Language in Parents’ Generics. On average, parents’ stereotype-incongruent generics included a greater proportion of hedging language (e.g., “Girls *can* play football!”; “Dads *sometimes* take care of babies”; *M* = 59% of utterances). However, parents used more definitive (non-hedging) language when generating stereotype-congruent generics (e.g., “Boys play football”, “Moms always fold the laundry”; *M* = 75% of utterances) and stereotype-neutral generics (e.g., “What do boys play?”, “Moms eat breakfast”; *M* = 71% of utterances).

Stereotypic Content of Parents’ Language. Parents’ production of stereotype-congruent ($\beta = 0.03$, *95% CI* [-0.11, 0.18]) and stereotype-incongruent talk ($\beta = 0.07$, *95% CI* [-0.08, 0.22]) remained stable across their children’s age (ps > .89).

Children’s Gender Essentialist and Stereotypical Beliefs

The following outlines task-specific analyses of children’s early gender essentialist and stereotypical beliefs (independent of their relation to parents’ language). We include task-specific findings for measures of children’s essentialist thinking (beliefs

about innateness, stability), as well as the measure of children's (own- versus other-) gender stereotypical thinking and preferences, and describe effects of child age, child gender, and measure-specific variables (e.g., stimulus gender) where appropriate. For details on our analytic approach, please refer to the main text.

Results

Innateness of Gender Category-related Features. Children endorsed essentialist beliefs about the innateness of gender category-linked properties approximately 69% of the time (*95% CI* [0.65, 0.73]). However, children demonstrated an own-gender bias in their essentialist thinking about the innateness of category-linked properties: girl participants endorsed greater essentialist beliefs about girls (M = 0.80, *95% CI* [0.73, 0.87]) compared to boys (M = 0.67, *95% CI* [0.59, 0.74]; contrast p = .01), and boy participants endorsed greater essentialist beliefs about boys (M = 0.76, *95% CI* [0.67, 0.85]) relative to girls (M = 0.51, *95% CI* [0.41, 0.61]; contrast $p < .001$; interaction of child gender and stimulus gender, Wald X^2 (1) = 17.98, $p < .001$). There were no changes in children's essentialist beliefs about the innateness of category-linked properties across the developmental window that we captured (p = .76).

Stability of Gender Category Membership. Children endorsed essentialist beliefs about the stability of category membership approximately 91% of the time (*95% CI* [0.90, 0.93]), and there were no effects of child gender or developmental time on children's beliefs about the stability of category membership (ps > .19).

Gender-Typed Preferences Task. Overall, children expressed greater interest in own-gender-typed (M = 2.55, *95% CI* [2.48, 2.62]) compared to other-gender activities (M = 2.09, *95% CI* [2.01, 2.17]; main effect of gender-type of property, Wald X^2 (1) =

88.54, $p < .001$). Additionally, children's interest in other-gender typed activities was greater among girls ($M = 2.20$, *95% CI* [2.11, 2.30]) compared to boys ($M = 1.93$, *95% CI* [1.80, 2.05]; main effect of child gender, Wald X^2 (1) = 5.74, $p = .02$) and decreased across child age for both girls and boys ($\beta = -0.62$, *95% CI* [-0.95, -0.28]; main effect of child mean-centered age, Wald X^2 (1) = 14.11, $p < .001$), whereas participants' interest in own-gender typed activities remained stable across child gender and age (ps > .13).

How does parental language relate to children's essentialist beliefs?

Stereotypic Content of Parent's Speech

The content of parents' language was generally unrelated to either dimension of their children's gender essentialist thinking. There were no main or interactive effects of either parents' stereotype-congruent or -incongruent talk on their children's essentialist beliefs about the innateness of category-related features (stereotype-congruent, β = -0.39, *95% CI* [-3.69, 4.46], ps > .84; stereotype-incongruent, $\beta = -2.13$, *95% CI* [-6.42, 2.15], ps > .34), or the stability of gender category membership across time (stereotype-congruent, $\beta = 0.69$, *95% CI* [-5.1, 6.48], ps > .53; stereotype-incongruent, β = -1.91, *95% CI* [-7.76, 3.95], ps > .52). There were no interactive effects of parents' stereotype-congruent or -incongruent language and developmental time (study session) on either of the gender essentialism measures (ps > .18).

Stereotypic Content of Parent Generics and Children's Composite Essentialism

In the main text, we briefly describe how parents' use of stereotype-congruent (e.g., "Boys don't cry"), -incongruent (e.g., "Girls are good at soccer"), and -neutral (e.g., "Moms eat breakfast") generics related to children's composite essentialist beliefs. Here,

we provide full statistics on those analyses: Children's composite essentialist beliefs were predicted by their parents' production of stereotype-incongruent generics (e.g., "Girls play football!"; β = 8.04, *95% CI* [2.82, 13.3]; main effect of stereotype-incongruent generics, Wald $X^2(1) = 7.73$, $p = .005$) and stereotype-neutral generics (e.g., "Moms eat breakfast", β = 1.36, *95% CI* [0.304, 2.42]; main effect of stereotype-neutral generics, Wald $X^2(1) = 6.29$, $p = .01$). Parental production of stereotype-congruent generics (e.g., "Boys don't wear nail polish!") were directionally consistent with this pattern, though the slope for stereotype-congruent generics did not reach significance on its own (β = 4.04, *95% CI* [-0.72, 8.8]; Wald $X^2(1) = 2.76$, $p = .09$).

How does parental language relate to children's gender stereotypical thinking?

Stereotypic Content of Parents' Generics and Children's (Other-Gender) Stereotypes

In the main text, we briefly describe how parents' use of stereotype-congruent (e.g., "Boys don't cry"), -incongruent (e.g., "Girls are good at soccer"), and -neutral (e.g., "Moms eat breakfast") generics related to children's other-gender stereotypes. Here, we provide full statistics on those analyses: Children's other-gender stereotypes were predicted by their parents' production of stereotype-congruent generics (e.g., "Boys don't wear nail polish!"; β = -7.96, *95% CI* [-14.2, -1.73], main effect of stereotype-congruent generics, $X^2(1) = 5.76$, $p = .02$) and stereotype-neutral generics (e.g., "Moms eat breakfast", β = -1.63, *95% CI* [-3.28, 0.03], main effect of stereotype-neutral generics, $X^2(1) = 3.86$, $p = .049$). Parental production of stereotype-incongruent generics (e.g., "Girls play football!") were directionally consistent

with this pattern, but did not reach significance on their own (β = -5.26, *95% CI* [-12.6, 2.03], $X^2(1) = 2.39$, $p = .12$).

Supplementary Analyses: Absolute Frequencies of Parents' Generics and Children's Gender Essentialist Thinking

To supplement our main analyses–which examined parents' relative use of generic language as a proportion of their total gender-marked talk–we also tested models using the absolute frequency (counts) of parents' generics as a predictor in analyses of children's essentialist thinking (innateness, stability). Whereas the proportional approach captures how generics are distributed within parents' overall speech, the absolute frequency approach reflects the total quantity of generic input a child receives, regardless of how much speech their parents produce overall.

In contrast to the proportionate use of generics in parents' gender-marked speech (see main text), the absolute frequency of parents' gendered generics use during the picturebook reading task was unrelated to their children's essentialist thinking about either the innateness of gender category-related features (β = 0.006, 95% CI [-0.03, 0.04]; ps > .71) or the stability of gender category membership (β = 0.05, 95% CI [-0.006, 0.10]; ps > .16). By comparison, analyses of the absolute frequency of parents' specific gender label usage were consistent with patterns observed in the proportion-based analyses. That is, parents who used a greater absolute number of specific gender labels had children who endorsed *less* essentialist beliefs about the innateness of gender category-related features (β = -0.027, 95% CI [-0.045, -0.009]; main effect of parent specific labels, Wald $X^2(1) = 9.16$, $p = .002$) as well as the stability

of gender category membership (β = -0.03, 95% CI [-0.06, -0.001]; interaction of session and parent specific labels, Wald $X^2(1) = 5.61$, $p = .017$), though this contrast did not reach significance in wave 2 of data collection (β = 0.01, 95% CI [-0.02, 0.05]).

As in the proportion-based analyses of the stereotypic content of generics reported in the main text, we also conducted exploratory analyses on the absolute frequencies of generics categorized as either counter-stereotypic (e.g., “Boys can play princess!”, M = 1.08), stereotypic (e.g., “Girls wear makeup”, M = 0.82), or neutral with respect to cultural gender stereotypes (e.g., “Moms eat breakfast”, M = 6.46). To maintain sufficient statistical power, we again focused on the composite measure of children’s gender essentialist thinking (collapsing across innateness and stability measures). Consistent with the proportion-based findings reported in the main text, analyses revealed that the absolute frequency of counter-stereotypic generics predicted greater composite child essentialism (β = .19, 95% CI [0.03, 0.35]; main effect of parents’ counter-stereotypic generics, Wald $X^2(1) = 4.32$, $p = .037$), with stereotypic and neutral generics showing directionally similar but non-significant trends (ps > .23). Taken together, these exploratory analyses suggest that children’s essentialist beliefs may be shaped not only by relative proportion of generics in parents’ gender-marked talk, but also by the absolute number of generic statements they explicitly hear, potentially regardless of whether those generics affirm, counter, or remain neutral with respect to gender stereotypes.

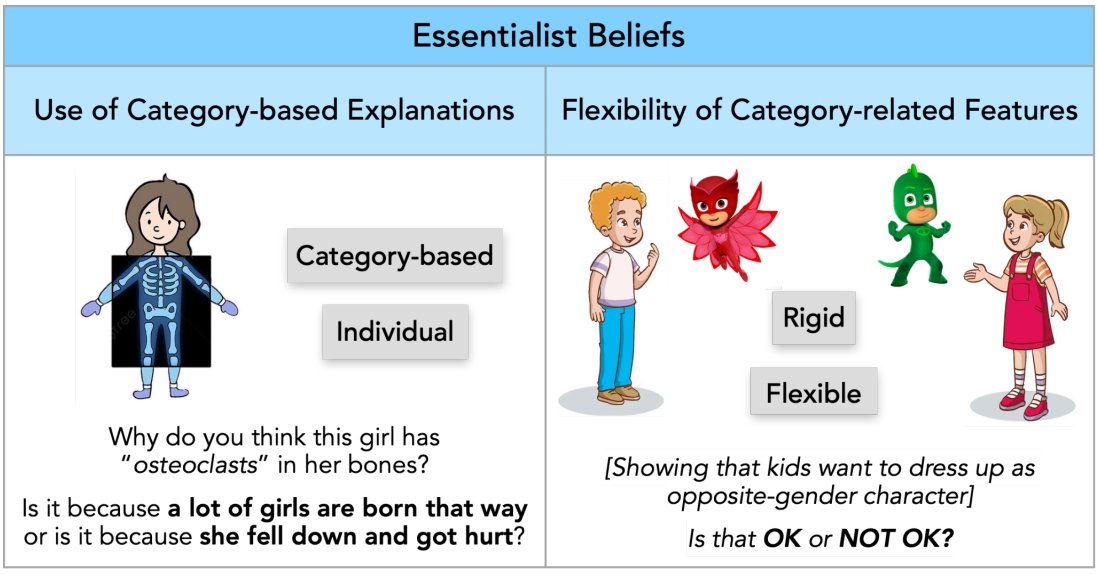
Additional Child Essentialism Measures

In the main manuscript text, we describe findings for those measures that

contributed to the essentialism composite (innateness, stability); here, we provide methods and results (independent of parent language) for additional measures of children's gender essentialist thinking (use of category-based reasoning, beliefs about flexibility of category-related features). These additional essentialism measures were not incorporated into the essentialism composite for several reasons. First, unlike the innateness and stability tasks that are standard in the essentialism literature (Gelman et al., 2004; Meyer & Gelman, 2016; Taylor, 1996), the following two new measures (category-based reasoning, flexibility beliefs) were adapted from research with an older age range, and produced unexpected and systemic methodological issues that did not allow for clear interpretation of their findings. For the category-based reasoning task, responses showed unexpected variation across property-type (i.e., social versus biological properties), and these effects further interacted with participant gender. For the flexibility task, a large subset of child participants were excluded for skipping critical questions altogether or for having non-stereotypical initial expectations (e.g., predicting that the girl would want to dress up as the masculine superhero) that made their responses difficult to interpret within the context of what this task was initially designed to assess. This pattern of findings was particularly unexpected given that this measure produced nearly unanimous stereotypical expectations when administered in a sample of slightly older children (e.g., five-year-olds; Foster-Hanson et al., 2022). For these reasons, we decided not to consider findings from these measures further, but we describe what they entailed and how they patterned across child age, child gender, and relevant measure-specific variables (e.g., property-type) for interested readers.

Methods

Gender Category-based Reasoning Task (New Measure). Participants were presented with four vignettes about boys and girls either exhibiting novel biological properties (e.g., “There’s a girl who has these things called ‘osteoclasts’ in her bones”) or engaged in novel social practices (e.g., “A lot of boys like wearing naggles on their hands”; see SOM Figure 1). For the biological items, children were asked whether the property originated from an internal category-based reason (e.g., “Because a lot of girls are born that way”; coded as “1”) or from an external individual reason (e.g., “Because she fell down and got hurt”; coded as “0”). For the social items, children were asked whether the social practice stemmed from an innate category-based reason (e.g., “Because that’s how a lot of boys’ hands are made”; coded as”1”) or from the result of cultural transmission (e.g., “Because a lot of boys learn how to wear naggles at school”; coded as “0”). Presentation of these items was counterbalanced by stimulus gender (boy, girl) and feature-type (behavioral, social) and randomized across participants; stimulus gender and feature-type were also included as predictors in these analyses.



SOM Figure 1. Summary of additional essentialism measures examining children’s use

of category-based explanations (panel 1, *n* = 157) and beliefs about the flexibility of category-related features (panel 2, *n* = 141).

Flexibility of Gender Category-related Features Task (New Measure). Participants were administered a gender flexibility task (Foster-Hanson et al., 2022), where they heard a story about a boy and girl who wanted to dress up as (stereotypically masculine and feminine) superhero characters, and were then asked which superhero they expected each child to dress up as (e.g., "Who should this kid be?"; see SOM Figure 1). Afterward, children were told that the boy wanted to dress up as the feminine character and the girl wanted to dress up as the masculine character. Children were asked to evaluate whether they thought the characters' decisions were "OK" (coded as "0", indicating a non-essentialist belief about the flexibility of gender category-related features) or "not OK," (coded as "1", indicating an essentialist belief about the flexibility of gender category-related features). Given that the purpose of this task was to probe children's normative judgments about the flexibility of stereotypically gender category-related features, individual test sessions from children who initially had non-stereotypical expectations (e.g., predicting that the girl would choose to dress up as the masculine superhero) were excluded from these analyses (*n* = 30). Additional child participant test sessions were excluded for not answering the expectation questions (*n* = 5), and skipping the test question (*n* = 1).

Results

Gender Category-Based Reasoning Task. Overall, children endorsed category-based explanations 43% of the time (*95% CI* [0.40, 0.47]), but these patterns differed by child gender, such that boy participants (*M* = 0.50, *95% CI* [0.45, 0.55])

engaged in more category-based reasoning than girl participants (*M* = 0.39, *95% CI* [0.35, 0.43]; main effect of child gender, Wald X^2 (1) = 4.94, *p* = .03). In addition, children (boys and girls) endorsed more category-based explanations when reasoning about biological (*M* = 0.46, *95% CI* [0.42, 0.51]) compared to social features (*M* = 0.40, *95% CI* [0.36, 0.45]; main effect of feature-type, Wald X^2 (1) = 4.92, *p* = .03), but this pattern varied by the gender of the stimulus (interaction of feature-type and stimulus gender, Wald X^2 (1) = 6.44, *p* = .01), such that on the biological items, children used more category-based reasoning when thinking about boy properties (*M* = 0.53, *95% CI* [0.47, 0.59]) compared to girl properties (*M* = 0.40, *95% CI* [0.34, 0.46]; contrast, *p* < .001), but on the social items, children endorsed category-based explanations at similar rates for both boy properties (*M* = 0.40, *95% CI* [0.34, 0.46]) and girl properties (*M* = 0.41, *95% CI* [0.35, 0.47]; contrast, *p* = .59). There were no significant changes in children's use of category-based explanations across developmental time (*p* = .94).

Flexibility of Gender Category-related Features Task. Children endorsed essentialist beliefs about the flexibility of gender category-related features 53% of the time (*95% CI* [0.46, 0.60])*;* there were no effects of child gender or developmental time on children's beliefs about the stability of category membership (*p*s > .40).

SOM References

Adolph, K., Tamis-LeMonda, C., Gilmore, R.O. & Soska, K. (2019). Play and Learning Across a Year (PLAY) Project - Protocols & Documentation. Databrary.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Brooks, M., Bolker, B., Kristensen, K., Maechler, M., Magnusson, A., & McGillycuddy, M. (2023). Package 'glmmtmb'. R Package Version, 1(1), 7.

Datavyu Team (2014). Datavyu: A Video Coding Tool. Databrary Project, New York University. URL http://datavyu.org.

Foster-Hanson, E., Leslie, S. J., & Rhodes, M. (2022). Speaking of kinds: How correcting generic statements can shape children's concepts. *Cognitive Science*, 46(12), e13223. https://doi.org/10.1111/cogs.13223.

Gelman, S. A., Taylor, M. G., & Nguyen, S. P. (2004). Mother-child conversations about gender: Understanding the acquisition of essentialist beliefs: I. Introduction. *Monographs of the Society for Research in Child Development.* https://www.jstor.org/stable/3701396

Meyer, M., & Gelman, S. A. (2016). Gender essentialism in children and parents: Implications for the development of gender stereotyping and gender-typed preferences. *Sex Roles*, 75, 409-421. https://doi.org/10.1007/s11199-016-0646-6

Taylor, M. G. (1996). The development of children's beliefs about social and biological aspects of gender differences. *Child Development*, 67(4), 1555-1571. https://doi.org/10.1111/j.1467-8624.1996.tb01814.x