Received: 15 November 2021 | Revised: 4 October 2023 | Accepted: 6 October 2023

DOI: 10.1002/icd.2475

COMMENTARY

WILEY

Examining the female-talker default in experimental language acquisition research

Annie Holtz[1] | Brandon T. Papineau[2]

[1]School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK

[2]Department of Linguistics, Stanford University, Stanford, California, USA

Correspondence
Annie Holtz, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD, UK.
Email: annie.holtz@ed.ac.uk

Abstract

Experimental research on language acquisition and development regularly employs auditory stimuli as part of the methodology. This project analyses the apparent standard practice of using female speakers to produce these experimental materials and the potential consequences of such a practice. To situate the discussion in the current scientific landscape we present a systematic review of published literature between 2017 and 2022 to establish how prevalent this practice is. The review finds a strong bias in favour of female-spoken stimuli across publications in a curated set of nine journals. We discuss this result in light of gender-based workplace inequality, changing caregiver expectations and the reliability of infants' assumed female voice preference. This project seeks to encourage researchers to consider how diversifying the stimuli used in these types of studies would lead to both a more inclusive and representative research landscape, as well as ensure that our research results are generalizable.

KEYWORDS
experimental stimuli, language acquisition, language and gender, language development, meta-science

1 | INTRODUCTION

There is a wealth of research in the developmental sciences that focuses on the interaction between the primary caregiver(s) and the child. Within the field of language acquisition, this focus was the foundation for coining the phrase 'motherese', as many studies explored the properties and effects of the type of speech that the assumed

default primary caregiver (the mother) uses when speaking to their child (Fernald, 1985; Gleitman et al., 1984; Newport, 1975). This speech register was defined by certain characteristics, such as an increase and/or expansion of the speech frequency range, short, one-clause sentences, repetition and an exaggeration of the phrasal intonation limits, and talking about objects and people in the immediate environment (Fernald & Kuhl, 1987; Papoušek et al., 1990; Snow & Ferguson, 1997). More recently, other terms have replaced 'motherese', as further research revealed that the use of this speech register extends to many other individuals who interact with children. These newer terms include 'parentese' (Clark, 2003; Ramírez-Esparza et al., 2017), 'caregiver/caretaker speech' (Klink & Klink, 1990; Ochs et al., 1979; Schachter et al., 1976), 'baby talk' (Morikawa et al., 1988), 'infant/child directed speech' (Clark, 2003; Ota et al., 2018; Rowe, 2012; Werker & McLeod, 1989) and so on.

The introduction of these terms marks a shift in focus for language acquisition research to also include in-depth studies of fathers' speech to their children (e.g. Benders et al., 2021; Niwano & Sugai, 2003; Pancsofar et al., 2010; Pancsofar & Vernon-Feagans, 2006) and, in the sociological literature, an increased focus on the changing nature and status of fatherhood in many cultures (e.g. Goldstein-Gidoni, 2020; Ranson, 2012; Seward & Rush, 2015; Van den Berg et al., 2013). Whilst these changes in nomenclature have shifted us towards a more inclusive and representative scientific landscape, the remaining traces of the mother-centric ideology in the current research domain need to be examined and their consequences explicated. In this paper, we focus on an example of these lingering ideologies. Namely, we explore the default practice of using female talkers in the creation of auditory stimuli in experimental child language acquisition research.[1] We believe that examining this practice constitutes an important exercise in critical engagement with long-standing methodological practices that may not have clear scientific motivation, but which may have emerged as a consequence of convenience or assumptions.

Our article has three main aims (i) to situate the discussion regarding the female speaker default in current practice, using a systematic literature review of articles in nine journals published over the last 5 years, (ii) to discuss the social and scientific implications of this practice for developmental research, with particular focus on research on language acquisition and (iii) to provide actionable solutions to some of the problems we highlight, as well as extend the call of 'design with diversity in mind' (Yarkoni, 2022) to also include considerations of the social structure in the country where the research is taking place.

2 | ORIGINS OF THE 'MOTHERESE' BIAS

Before examining the prevalence of the use of female voices in contemporary research on language acquisition, it is worth highlighting the motivations behind this practice. There are a number of such motivations, both practical and theoretical. We touch on three of them here, before returning to these motivations in Section 4, in light of our systematic literature review findings.

The first of these motivations is practical in nature: the gender division amongst the scientists working in developmental research areas (such as developmental psychology) is such that these fields tend to include more female than male researchers (Fletcher et al., 2010; Fowler et al., 2018). As such, when finding potential speakers who can record stimuli for their experiments, many researchers turn first to their colleagues and collaborators. This 'stimuli of convenience' results in a high number of studies where female voices are used in the creation of the experimental stimuli.

The second motivation is based on the assumption that the vast majority of children are primarily exposed to female voices in their day-to-day lives, as they receive input in both home and child-care settings, which are strongly associated with domestic femininity in popular discourse. To create maximally naturalistic experimental settings, then, researchers take into account the assumed input of the children they test in the lab. The result of this is a tendency to employ the voices of women in the experimental setting.

The last motivation, and perhaps the most interesting one, stems from research showing that children, especially infants, have a preference for particular voices. For example, infants prefer spoken linguistic stimuli provided by their own mother over stimuli spoken by unknown female speakers (DeCasper & Fifer, 1980), whereas no similar

preference is found in favour of their father's voice over voices of unknown male speakers (DeCasper & Prescott, 1984). This also generalises to infants' vocal response to infant-directed speech (hereafter IDS), where they respond more to IDS spoken by their mother, compared to their father (Niwano & Sugai, 2003). Similar evidence is also found in neuroimaging studies examining levels of brain activation in infants, wherein more neural activity is observed in infants' brains when listening to female-spoken IDS compared to male-spoken IDS (Sulpizio et al., 2018). The result of these findings is that many researchers have adopted female voices in their experiments with the intention that this will keep young participants engaged in the specific task at hand. This course of action is in turn intended to mitigate the traditionally high drop-out rates and data loss in experiments involving children, due to inattention or fussiness (Slaughter & Suddendorf, 2007; Stets et al., 2012; Van der Velde & Junge, 2020).

Having now reviewed some of the reasons that the practice of employing female voices in experimental child language research has become the standard in the field, we turn to an empirical investigation of this practice in recent years. Such an examination allows us to highlight the starkness of this practice before turning to the consequences of it.

3 | FEMALE VOICES IN EXPERIMENTAL LITERATURE

To confirm the presence of this practice, and to quantify the extent to which the practice is present in the current research landscape, we conducted a systematic literature review of experimental research on language acquisition in a selection of nine journals in the last 5 years. In this section, we provide an overview of the methods and results of the systematic literature review. Full details of the search criteria employed can be found in the preregistration. A link to the preregistration is provided in the Data Availability Statement.

3.1 | Search strategy and selection

The systematic review of the literature was conducted by manually searching a specific set of journals, selected for their focus on language acquisition, language development and/or developmental psychology. The journals were selected by identifying important publications in these three areas using impact factor searches and in consultation with researchers working in these same areas. The selected journals were *Language Acquisition*, *Language Learning and Development*, *Journal of Child Language*, *Child Development*, *Infant and Child Development*, *Developmental Psychology*, *Developmental Science*, *Infancy* and *Infant Behavior and Development*. The search included all papers published between 2017 and 2022 (inclusive).[2] Studies in the aforementioned journals were evaluated for inclusion based on the following criteria.

1. Stimuli criteria
   a. Only experimental studies that employ prerecorded vocal stimuli were included.
   b. Experimental studies that employ *only* synthesised voices were excluded; if at least one natural speaker was used, the paper was deemed to satisfy this criterion. We accepted stimuli which were artificially manipulated in the creation process, to include studies that employed phonetic or lexical splicing, so long as the original stimulus retained its natural status.
   c. Experimental studies in which the audio data are part of a video recording were included.
   d. Experimental studies that used the parents' or caretakers' voice(s) were included, provided that the stimuli were prerecorded.
   e. Experimental studies could employ any target language (encompassing both artificial and natural languages, as well as language tokens such as syllables), as long as it included spoken language.

f. Studies that did not take place in a lab or equivalent environments were excluded.[3]

2. Other criteria
   a. We included research published between the dates 1/1/2017 and 19/4/2022 (the latter being the submission date of the project preregistration, available here).
   b. Studies where all participants were above the age of 5 were excluded.
   c. Studies that did not focus on some aspect of language acquisition or development were excluded (e.g. we excluded studies that examined anxiety disorders in children but which happened to use audio stimuli in their procedure).
   d. Studies could be behavioural or neurophysiological, as long as the main research question concerned language acquisition/development.

Following Ahn and Kang (2018), who outline good research practices for conducting literature searches to use as the basis of a systematic review, both authors conducted the literature search independently using these pre-determined inclusion/exclusion criteria. Each author curated a list of information about the papers that meet the inclusion criteria based on a predetermined template, which can be found in the OSF repository linked in the Data Availability Statement. The information gathered included the name of the journal the article was published in, year of publication, volume and issue number, author names, country or countries where the research was conducted and so on.[4] After the full set of literature was searched by both authors, the two lists were compared and any studies that only appeared in one of the two lists were revisited jointly by both authors and their inclusion/exclusion was discussed. Once agreement had been reached, the full list of papers comprised 273 studies.

3.2 | Speaker gender coding

Each author coded half of the agreed-upon articles according to the stimuli speakers' reported gender/sex. Reports of either speakers' gender or sex were taken as equivalent for the purposes of data collection and we accepted both direct statements of gender (e.g. 'a female speaker'), as well as pronominal anaphora as evidence of speaker gender (e.g. 'she recorded three sentences…'). When no gender was included in the description of the stimuli, the first author emailed the paper's corresponding author to inquire after this. The 80% of papers stated the gender of the stimulus speaker explicitly, and a further 8.75% were obtained from emails sent to article authors. The 7.3% of papers made no indication of speaker gender, and the remaining papers were coded by pronominal anaphora or because the speaker was explicitly stated to be an author.

The gender coding scheme was quinary in nature, distinguishing between Male (M), Female (F), Non-binary (NB)[5], Mixed (Mix) and Unknown (U) stimuli speaker genders. We then randomly sampled 10% of the papers coded by each author and had them re-coded by the other author. Based on the agreement data from this comparison we computed an unweighted Cohen's Kappa using the *vcd* (Meyer et al., 2022) package in R, yielding an agreement score of 0.84 (CI: 0.63–1) which falls within the range of near perfect/strong agreement between coders (Landis & Koch, 1977; McHugh, 2012). Only 2 instances of disagreement were found between coders and these were resolved by taking a second pass at the specific papers in question by both authors jointly. Following this, the remainder of the coded data from both authors were combined to form the final fully-coded data file.[6]

3.3 | Results

To better situate our findings, we note that the majority of our papers came from 3 journals: *Developmental Science*, *Child Language* and *Infancy*. These three journals alone accounted for over 56% of the included papers. The fewest

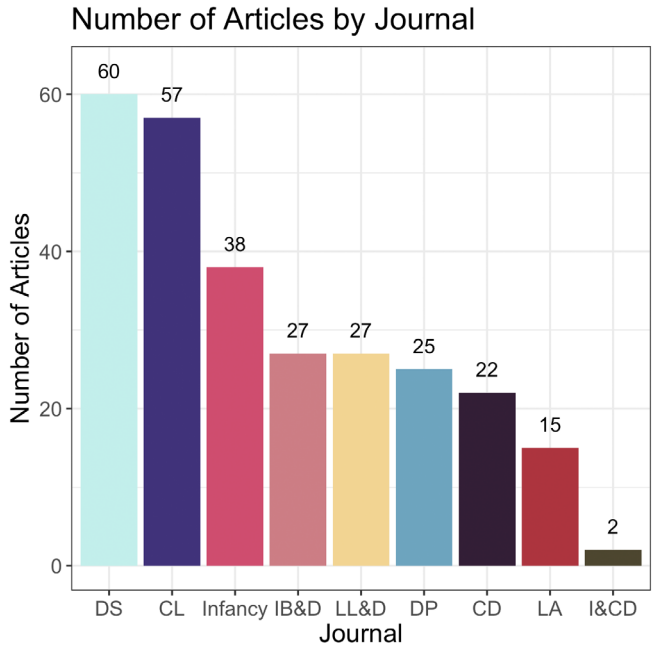

FIGURE 1 Included journals by number of included papers from each source. CD, child development; CL, child language; DP, developmental psychology; DS, developmental science; I&CD, infant and child development; IB & D, infant behaviour and development; LA, language acquisition; LL&D, language learning and development.

papers came from *Infant and Child Development*, with only two. The rest came from the remaining five journals. These numbers are reported in Figure 1.

The results of our literature review paint a stark picture that confirms the aforementioned bias towards using female voices in experimental child language acquisition research. Of the 273 papers coded in the course of our literature review, 215 used only female talkers in stimuli creation – a whopping 78.7%. Only 38 papers included male voices at all; 26 used a mix of male and female voices and 12 used only male voices.[7] The remaining 20 papers did not indicate the gender of their talkers and did not respond to the first author's emails on this question within the preregistered time frame. The results of this search are presented visually in Figure 2.

4 | DISCUSSION

The results of the previous section's literature review served to validate the assumption that the stimuli employed in experimental child language acquisition studies are dominated by female talkers. Before we discuss potential issues that arise as a result of this practice, we provide an outline of those few studies which used either only male voices or a mixture of male and female voices.

Of the 12 studies that included only male voices, none gave specific scientific motivation for this design choice; rather it was reported in a manner similar to those papers that employ only female voices – that is to say without motivation. None of the studies examined a research question which, in and of itself, meant that using a male voice was highly motivated.

The clearer explanation for speaker choice was generally provided in the 26 papers which used a mixture of female and male voices. Several of these papers explicitly stated that speaker gender was used as a cue for different speakers (e.g. Ferguson et al., 2018). In these studies, the research question often involved the perception or

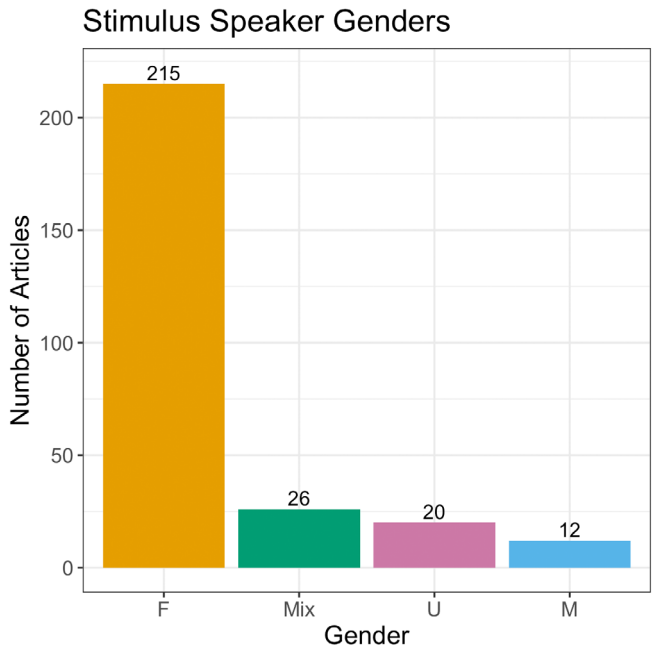

FIGURE 2 Results of literature review by gender of auditory stimuli employed.

detection of language features/speaker properties in diverse contexts (e.g. Fecher & Johnson, 2019) and so using speakers of different genders allows for testing of children's ability to generalise across input sources (Siying & Renji, 2019; ter Haar & Levelt, 2018). A small number of studies alternated speaker genders as a way of differentiating between the training and test stages of the experiments (e.g. Bulgarelli & Bergelson, 2021; Verdine et al., 2017).

Since the manipulation of speaker gender was generally not the main research question for these studies (the exception being Sulpizio et al. (2018), which found differences in neural activation in response to male and female-spoken IDS), few provide breakdowns of their results by speaker gender. Those that do, such as Singh's (2018) examination of cross-speaker word recognition in infants, reported no difference in test performance between groups that were trained on male and female voices. The use of male voices thus seems to be either an unmotivated design choice or incorporated as a useful cue for differentiation within experimental designs whose primary research questions do not involve gender differences.

Returning to the overall default use of female voices, we now discuss potential issues that arise as a result of this practice, presenting them as counterpoints to the motivations outlined in Section 2. We begin with a discussion of gender inequalities in the academic workplace before turning to the issue of motherhood as the default model of parenthood. We conclude with a discussion of the accepted notion that infants prefer female voices over male ones in light of the present body of evidence on this issue.

4.1 | Gender inequalities in the academic workplace

Research has found that female members of academic staff tend to perform a larger portion of the pastoral and internal services directed at supporting the department they work in and the students that study there (Babcock et al., 2017; Guarino & Borden, 2017; Misra et al., 2011; O'Meara et al., 2017). Furthermore, female professors are more likely to receive requests for special favours or concessions from students than their male colleagues (El-Alayli et al., 2018), resulting in either an increase in work to accommodate such requests or managing the effects of

denying the students such requests. As universities increasingly focus on trying to attract members of underrepresented groups into the academic workforce and student body, women and non-binary individuals (especially those of minority ethnic backgrounds) are more actively involved in diversity and inclusion initiatives than their male peers (Jimenez et al., 2019). This service workload impacts their productivity in other areas of their work, such as research output, which tends to receive higher weighting during considerations for promotion than services to the department and wider university community (Alperin et al., 2019). Devaluation of the service work performed by women is evident in the attitudes of academics themselves, as they tend to evaluate service tasks performed by men as more valuable than those performed by women (Hanasono et al., 2019).

As one reviewer pointed out, it is possible that the gender imbalance seen in the data we have collected is due to authors using their own voices, and there is a documented imbalance in the graduate student body of developmental psychology, such that, for example, the majority of PhD recipients in the United States in any given year is majority female (Office APAWP, 2006). However, a 2015 American Psychological Association report notes that, whilst the proportion of women in faculty positions is increasing in psychology at large, the proportion of women is currently hovering around 50 – much lower than our findings would predict if authors were using their own voices. The National Science Foundation similarly found that women accounted for 69% of early career academics in psychology (Early Career Doctorates Survey, 2017) and 58% of psychology post-docs (Survey of Graduate Students and Postdoctorates in Science and Engineering, 2018). Turning to the United Kingdom we find that 62% of lower-level faculty staff are female (Lai et al., 2022, based on data from 38 UK universities). This shows that, although women are in the majority in this academic group, the proportion of female-only stimuli papers is still higher than the number of female early-careers academic psychologists in the United Kingdom. Whilst we did not code for author gender, it remains our impression that both male and female authors reported using female-speaker stimuli. Additionally, it is worth noting that the papers examined herein were authored not only by developmental psychologists but also by linguists. Whilst the numbers for linguistics are somewhat harder to attain, grassroots attempts at documenting gender imbalances have found a slight favouring of women in the graduate student body, but a significant majority of full professors are male (Dickerson et al., 2019).

The use of a default speaker in a given lab, often recruited from within the department, compounds previously identified issues regarding the division of labour within the academic workforce. As researchers who wish to use recorded voices in their experiments need to recruit speakers, female members of staff or research students are often recruited for this purpose. This work, as with much of the other pastoral labour conducted by female academics, is often unpaid and may constitute a longer time-commitment if retakes or follow-up studies need to be conducted. If a particular woman amongst the staff is found to have an exceptionally 'good' voice for this type of research, she may end up receiving many requests to record stimuli and subsequently find it hard to decline such requests as she has performed the service before. Using female-only stimuli in experiments the academic community places an additional undue burden of expectation on female staff and students to perform unpaid labour.

One way to mitigate the effects of this practice is to make it standard across institutions to pay a fair hourly wage for this work. Doing so would increase the observed value of this type of work by acknowledging that the time spent performing such tasks is worthy of payment. This is especially pertinent in a scientific landscape where it is increasingly common to pay participants in lab and online experiments a fair wage and/or substantial incentive for participating. For example, the popular crowdsourcing platform Prolific requires researchers using their platform to pay all participants a fair hourly wage (Prolific Team, 2022). Whilst the time and effort these women spend on stimuli recording may seem negligible in comparison to the time they spend on other service tasks that are also unpaid, it is part of the aforementioned wider pattern of women performing unpaid or devalued services in academia.

4.2 | Motherhood as parenthood

The a priori assumption across the globe is that women are the primary caregivers of children, resulting in the assumption that children are primarily exposed to language produced by women. Indeed, this default seems to hold a

degree of truth in it. In the United States, for example, only about 17% of stay-at-home parents were fathers in 2016 (Livingston, 2018). Despite this glaring imbalance, however, there is evidence that the scales are tipping towards a more equal share of child-rearing. With regards to single parents, for example, in 2020, nearly 25% of all single-parent households in the United States consisted of only a father, compared to under 10% in the 1970s (U.S. Census Bureau, 2020), and a similar increase has been seen in Canada over recent decades (Statistics Canada, 2014). This is not universally the case, however, and the United Kingdom's stagnation of single fatherhood (which has hovered around 10% of all single-parent households for over two decades) presents a stark counterexample to this pattern (Office for National Statistics, UK Statistics Authority, 2021).

Whilst the trends so far reported have dealt only with Western, English-speaking countries, they are far from being isolated phenomena. Goldstein-Gidoni (2020), for example, elucidates the experiences of Japanese 'working fathers', or fathers who increasingly structure their work and social lives around family and fatherhood in contemporary Japan (Ranson, 2012). This trend is positioned as an emergent gender realisation, with men taking up domestic engagement than in bygone eras. Van den Berg et al. (2013) similarly report that the 'One Man Can' programme has positively impacted fathers in South Africa, such that more fathers in the region are spending more time with their children and engaging in domestic responsibilities and communications more regularly. These trends are reflective of investigations of fathering trends the world over, which have largely pointed to a growing involvement of fathers in child-rearing and domestic tasks; Seward and Rush (2015) provide a more comprehensive overview of the current state of fatherhood trends.

Finally, the fact of the matter remains that, despite significant legal and logistical roadblocks (Tornello et al., 2019), queer, trans and other gender non-conforming individuals are having and raising children (see Stotzer et al., 2014 for an overview of trans parenting trends, and Manning et al., 2014 for a review of child development studies involving same-sex parenting). The United Kingdom, for example, estimates that there were more than 2000 same-sex couples with children residing in the country in 2019, though no parent-gender breakdown is provided (Office for National Statistics, UK Statistics Authority, 2021). Similarly, Canada reported in 2011 that over 1900 children under the age of 24 were living with same-sex male parents, and trends in the Canadian census indicate that the numbers of same-sex marriages (both common-law and official) continue to grow rapidly, nearly trebling between the years 2006 and 2011 (Statistics Canada, 2015). Stotzer et al. (2014) estimate, based on a review of more than 50 published studies on transgender and gender non-conforming parenting that some 25%–50% of trans adults in the United States report being parents. Whilst these numbers are undoubtedly different now, almost a decade on, with increased recognition and protection for trans adults and parents, they indicate that a non-trivial number of children are being raised by parents to whom the traditional norms of gender might not apply.

Taken together, these figures all reflect a global parenting scene that looks very different from the parenting scene of 40 or 50 years ago. Increasingly, fathers and other male parental figures are becoming involved in child-raising, domestic tasks and single parenthood. In addition, the increasing protection and recognition of queer, trans and non-binary parents mean that an ever-growing number of children are being raised in households that challenge the traditional domestic configuration that places women and mothers in heterosexual relationships at the forefront of children's development.

These trends mean that the linguistic experiences and development of children raised by single fathers, stay-at-home fathers, same-sex parents and trans and non-binary parents have been understudied thus far. The issues introduced by this are two: in the first instance, it means that we as academics are erasing the lived experiences of the children who are raised by people other than women. We are also implicitly upholding a hetero and cis-normative standard of parenthood. Secondly, there is a closely-related issue of generalizability. By ignoring the experiences and data of traditionally marginalised groups, we are painting incomplete pictures of the issues we study. Such an oversight means that we may be missing pieces of the language acquisition puzzle that we had heretofore not considered. Whilst accounting for all possible family configurations and their effects on language acquisition is of

course unfeasible, we as researchers do have a duty to 'design with diversity in mind' (Yarkoni, 2022) and advocate for a more inclusive scientific landscape.

Whilst not defaulting to using only female voices in our stimuli is one way to help address this issue, another effective tool would be to consistently collect data about our participants' primary caregiver situation. This would allow us to include the effects of caregiver gender identity in statistical models, helping account for possible individual differences in participant behaviour based on their naturalistic language input.[8] In addition to diversifying our experimental design, this would allow us to also *analyse* with diversity in mind.

4.3 | Reliability of the female voice preference

Whilst there is some evidence in support of the notion that children show a general preference towards female voices, much of the evidence for this relates to a preference for the children's own mothers' voices, rather than female voices in general. Indeed, behavioural studies that contrast infant attentiveness to IDS spoken by an unknown female and male voices have failed to find differences that meet traditional standards of significance, such as in Werker and McLeod (1989; $p = 0.053$). The main difference was instead in affective response, where infants did show higher response rates to the stimuli produced by the female speaker (Werker & McLeod, 1989). Worth noting is that this study also included video footage of the speakers, so any attention and affective response differences may also be due to visual rather than auditory preferences. Whilst some recent investigations indicate a possible difference in infants' neural activation when exposed to male and female voices (Sulpizio et al., 2018), behavioural examinations of infants' ability to extract linguistic information from stimuli found no difference based on speaker gender (Houston & Jusczyk, 2000).

In addition, some of the research cited in support of the general female voice preference is highly ambiguous with regard to its methodology. For example, one paper cited in several of the studies we refer to above is Brazelton (1978), where the evidence in support of this preference is the following statement:

> The responses to auditory stimuli are similar, and demonstrate again the choice-making capability of the baby. If a male and a female stand on opposite sides of a newborn and talk at the same consistent rate, the newborn will stop moving, his face will knit, and he will turn towards the female's voice over and over and over. This test is now done with prematures to get some idea of integrity from this kind of choice making. [p. 188]

This claim in the paper is not accompanied by any citation or reference to specific evaluations, likely as a function of the paper being a report of a conference talk given by Brazelton. As such, it remains unclear how reliable the preference for female voices in general is for infants, and how long such a preference might persist during childhood. In addition, it is unknown if this preference generalizes to infants whose primary caregiver(s) is/are not their mother (or another woman), as this would entail that they are less habituated to female voices.

In general, using heavily restricted stimuli sets when exploring a phenomenon such as language acquisition restricts our ability as researchers to draw adequate generalisations based on those experimental results. This point is made quite strongly by Simons et al. (2017), who argue for the inclusion of a 'Constraints on Generality' statement in empirical papers, which would force authors to make explicit their reasons for believing, or not believing, that their findings are generalizable to a larger population than simply that of the sample. They do not limit these constraints on generalizability to just populations and participants, however, and specifically call for discussions of the role of stimuli and procedures as well. We join them in this call for explicit discussions of generalizability, along with many others in the fields of psycholinguistics and psychology (Baribault et al., 2018; Visser et al., 2022; Yarkoni, 2022, inter alia).

If we find that children are able to extract statistical dependencies between segments in a specific stream of synthesised speech, for example, then this needs to be replicated both with different participants and with a

different stream of sounds. This would enable us to say that those results generalise beyond 'children from this one area, and of this specific age group, are able to extract statistical dependencies from *this* stream of synthesized speech.' One dimension to consider in the possible configurations of 'different streams of sounds' is the gender dimension, and the inclusion of non-female-talker stimuli would augment our ability as researchers to make more general claims about the implications and validity of our findings.

Related to this issue of generalizability, and of particular relevance to the discussion of stimuli creation, is the fact that there is currently no standard practice in place for noting how many unique speakers are being used across studies on language acquisition. As noted in Section 4.1, it is not uncommon for the same woman to be approached to record stimuli several times and for different projects. Consequently, studies from a particular labour department may contain hidden dependencies based on speaker identity which could cause talker-specific effects that are masked when such studies are included in meta-analyses (Winter & Grice, 2021). This observation has already resulted in an initiative to use more varied voice stimuli to improve generalizability in experimental research and has been implemented in several studies outwith developmental research (e.g. Baumann & Winter, 2018; Cangemi et al., 2015; Roettger et al., 2014).

Whilst using a more varied selection of speakers both within and across studies helps mitigate some of the worries relating to generalizability, being able to identify and include speaker identity in meta-analyses would also be hugely beneficial for the robustness of research findings. One way of doing this would be to create a dedicated database where speakers register, potentially providing some other demographic details about themselves, and which generates an ID that could then be included in all papers where their voices serve as the stimuli. A possible downside, if this is an identifiable record of the speakers, is that those speakers who end up being associated with 'positive' results might end up receiving even more requests for stimuli recording, or stimuli from those speakers/studies will be selectively used to engineer more positive experimental results.[9] Ideally, this database would be anonymous, thus making it harder to identify specific speakers but still allowing for the use of their IDs during data analysis and meta-analyses. Whilst the creation of such a database is outwith the remit of this paper, we believe its creation and implementation would be of benefit to the research community at large.

In summary, the assumption that infants prefer women's voices requires further experimental testing to evaluate if this is a general preference or specifically related to a preference for their mother's/caregiver's voice. Further exploration of how long such a preference may persist and/or under what conditions it disappears would enable researchers to make an informed decision about how to craft stimuli that are engaging for their target population. Finally, being more transparent regarding our use of stimuli voices, and using a wider variety of voices whenever possible, will enable us to more adequately evaluate the generalizability of our research results. Whilst gender represents only one of many possible dimensions that could be varied during stimuli creation for experimental research to ensure true generalizability, it is an especially interesting one to examine at this time, due to the shifting nature of caretaker roles in many countries and the growing body of work examining properties of IDS/CDS across different configurations of caretakers.

5 | CONCLUSION

We have shown that the female voice holds a particularly privileged position in the realm of experimental child language research; the results of our literature review indicated that nearly 80% of the papers we surveyed employed only female voices. Whilst there are many good reasons for this practice, including the convenient proximity of women to child language studies, the tendency for women to be primary caregivers, and the purported preference of children for female voices, the practice also comes with a number of concerning drawbacks. We have addressed three of them here, highlighting the way the practice perpetuates gender inequalities and invisible labour in academia, as well as cisheteronormative family configurations. Finally, we have shown that question of children's' preferences for female voices merits further investigation; it remains unclear if the observed preference for female voices is in fact a result of speaker gender, or if it rather has to do with familiarity and speaker-level effects.

Whilst the work presented herein has argued that the current practice of treating the female voice as the default is problematic in nature, we conclude by stating our hope that rather than being interpreted as an admonishment, our elucidation of this phenomenon will open up new and exciting avenues of investigation. For example, the explicit investigation of more diverse family configurations in our research provides us with the opportunity to examine the potential effects of such configurational variability. Similarly, the inclusion of male and non-binary voices in our stimuli might lead to insightful discoveries about the way that gender-indexing features of language are parsed and processed by young children. We leave these questions for future investigation and offer the arguments herein as an impetus for that research to be carried out.

AUTHOR CONTRIBUTIONS

Annie Holtz: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; visualization; writing – original draft; writing – review and editing. **Brandon T. Papineau:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; visualization; writing – original draft; writing – review and editing.

ACKNOWLEDGEMENTS

We would like to extend our thanks to the editorial team at Infant and Child Development as well as the three anonymous reviewers of this paper, whose comments greatly helped shape this work. We also thank Jennifer Culbertson, Nina Markl, Mits Ota, Barbora Skarabela, and Niklas Wahlberg for their support and advice in the development of this project. Any remaining errors are our own.

CONFLICT OF INTEREST STATEMENT

The authors report no conflict of interest in the development or publication of this work.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/icd.2475.

DATA AVAILABILITY STATEMENT

The data collected for this study, along with all analysis codes, are openly available on OSF. https://osf.io/u4snt/. The associated preregistration can be found here: https://osf.io/ycs6e.

ORCID

Annie Holtz https://orcid.org/0000-0002-4325-9931

Brandon T. Papineau https://orcid.org/0000-0003-1298-0731

ENDNOTES

[1] We recognize that in referring to 'mothers' and 'fathers', we implicitly cast male and female gender categories as contrasting and binary in nature. We do not endorse the notion of the gender binary but adopt the language of 'mothers', 'fathers', 'female', and 'male' to keep with previous research.

[2] We elected to go by online publication dates (if this information was available) and so any study published on/after 1st January 2017 until 19th April 2022 were eligible for inclusion.

[3] We accepted quiet individual rooms in a school/nursery/the child's home for this criterion, but not other public places.

[4] Note that there are some differences between the template designed for the preregistration and the one used during data collection as we realised that some extra information was needed (e.g. volume and issue number) and some information was impractical to collect (e.g. the type of study, as the classification for this was unclear).

[5] No non-binary speakers were found in the sample of papers, so they are not included in the visualisations below.

[6] Note that during the revision process a further 3 incorrectly coded speaker gender classifications were discovered and corrected.

[7] The dataset originally included 274 papers, however one paper was excluded after an author confirmed that the study included no prerecorded stimuli.

[8] We thank an anonymous reviewer for highlighting this possibility to us.

[9] We thank an anonymous reviewer for highlighting this as a potential issue.

REFERENCES

Ahn, E., & Kang, H. (2018). Introduction to systematic review and meta-analysis. *Korean Journal of Anes Thesiology*, *71*(2), 103–112. https://doi.org/10.4097/kjae.2018.71.2.103

Alperin, J. P., Nieves, C. M., Schimanski, L. A., Fischman, G. E., Niles, M. T., & McKiernan, E. C. (2019). Meta-research: How significant are the public dimensions of faculty work in review, promotion and tenure documents? *eLife*, *8*, e42254.

Babcock, L., Recalde, M. P., Vesterlund, L., & Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, *107*(3), 714–747.

Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., Van Ravenzwaaij, D., White, C. N., De Boeck, P., & Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences*, *115*(11), 2607–2612.

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained german listeners' perceptual judgments. *Journal of Phonetics*, *70*, 20–38.

Benders, T., StGeorge, J., & Fletcher, R. (2021). Infant-directed speech by dutch fathers: Increased pitch variability within and across utterances. *Language Learning and Development*, *17*(3), 292–325.

Brazelton, T. B. (1978). The remarkable talents of the newborn. *Birth & the Family Journal*, *5*(4), 187–191.

Bulgarelli, F., & Bergelson, E. (2021). Talker variability shapes early word representations in english-learning 8-month-olds. https://doi.org/10.31234/osf.io/rxyjc

Cangemi, F., Krüger, M., & Grice, M. (2015). Listener-specific perception of speaker-specific production in intonation. In S. Fuchs, D. Pape, C. Petrone, & P. Perrier (Eds.), *Individual differences in speech production and perception* (pp. 123–145). Peter Lang.

Clark, E. V. (2003). *First language acquisition*. Cambridge University Press. https://doi.org/10.1017/CBO9781316534175

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

DeCasper, A. J., & Prescott, P. A. (1984). Human newborns' perception of male voices: Preference, discrimination, and reinforcing value. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *17*(5), 481–491. https://doi.org/10.1002/dev.420170506

Dickerson, B., Durvasula, K., & Liter, A. (2019). Bias in linguistics. https://biasinlinguistics.org/about/

El-Alayli, A., Hansen-Brown, A. A., & Ceynar, M. (2018). Dancing backwards in high heels: Female professors experience more work demands and special favor requests, particularly from academically entitled students. *Sex Roles*, *79*(3), 136–150. https://doi.org/10.1007/s11199-017-0872-6

Fecher, N., & Johnson, E. K. (2019). Bilingual infants excel at foreign-language talker recognition. *Developmental Science*, *22*(4), e12778.

Ferguson, B., Graf, E., & Waxman, S. R. (2018). When veps cry: Two-year-olds efficiently learn novel words from linguistic contexts alone. *Language Learning and Development*, *14*(1), 1–12.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

Fletcher, J., Bloor, K., Crossman, C., Thornton, J., Briggs, E., Hawkins, T., Sammut, S., & Cardwell, K. (2010). Profiling the college of educational and developmental psychologists: An examination of demographics, professional practice, attitudes and professional development preferences. *The Educational and Developmental Psychologist*, *27*(1), 1–19.

Fowler, G., Cope, C., Michalski, D., Christidis, P., Lin, L., & Conroy, J. (2018). Women outnumber men in psychology graduate programs. *Monitor on Psychology*, *49*(11), 21.

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language*, *11*(1), 43–79. https://doi.org/10.1017/S0305000900005584

Goldstein-Gidoni, O. (2020). 'Working fathers' in Japan: Leading a change in gender relations? *Gender, Work and Organization*, *27*(3), 362–378.

Guarino, C. M., & Borden, V. M. (2017). Faculty service loads and gender: Are women taking care of the academic family? *Research in Higher Education*, *58*(6), 672–694. https://doi.org/10.1007/s11162-017-9454-2

Hanasono, L. K., Broido, E. M., Yacobucci, M. M., Root, K. V., Peña, S., & O'Neil, D. A. (2019). Secret service: Revealing gender biases in the visibility and value of faculty service. *Journal of Diversity in Higher Education*, *12*(1), 85–98.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582. https://doi.org/10.1037/0096-1523.26.5.1570

Jimenez, M. F., Laverty, T. M., Bombaci, S. P., Wilkins, K., Bennett, D. E., & Pejchar, L. (2019). Under-represented faculty play a disproportionate role in advancing diversity and inclusion. *Nature Ecology & Evolution*, *3*(7), 1030–1033.

Klink, M., & Klink, W. (1990). The influence of father caretaker speech on early language development: A case study. *Early Child Development and Care*, *62*(1), 7–22. https://doi.org/10.1080/0300443900620102

Lai, K. A., Saxena, G., & Allen, P. J. (2022). Research performance of academic psychologists in the United Kingdom. *Scientometrics*, *127*(7), 4139–4166.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Livingston, G. (2018). *Stay-at-home moms and dads account for about one-in-five us parents*. Pew ResearchCenter.

Manning, W. D., Fettro, M. N., & Lamidi, E. (2014). Child well-being in same-sex parent families: Review of research prepared for american sociological association amicus brief. *Population Research and Policy Review*, *33*(4), 485–502.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Meyer, D., Zeileis, A., & Hornik, K. (2022). *Vcd: Visualizing categorical data* [R package version 1.4-10].

Misra, J., Lundquist, J. H., Holmes, E., Agiomavritis, S., et al. (2011). The ivory ceiling of service work. *Academe*, *97*(1), 22–26.

Morikawa, H., Shand, N., & Kosawa, Y. (1988). Maternal speech to prelingual infants in Japan and the United States: Relationships among functions, forms and referents. *Journal of Child Language*, *15*(2), 237–256. https://doi.org/10.1017/S0305000900012356

National Center for Science and Engineering Statistics (NCSES). (2020). *Survey of Graduate Students and Post doctorates in Science and Engineering, Fall 2018*. Alexandria, VA: National Science Foundation. http://ncsesdata.nsf.gov/gradpostdoc/.

National Center for Science and Engineering Statistics (NCSES). (2021). *Early Career Doctorates: 2017*. NSF 21–323. Alexandria, VA: National Science Foundation. https://ncses.nsf.gov/pubs/nsf21323/.

Newport, E. L. (1975). *Motherese: The speech of mothers to young children*. (Doctoral dissertation). University of Pennsylvania. https://repository.upenn.edu/dissertations/AAI7524108

Niwano, K., & Sugai, K. (2003). Pitch characteristics of speech during mother-infant and father-infant vocal interactions. *The Japanese Journal of Special Education*, *40*(6), 663–674.

Ochs, E., Schieffelin, B. B., & Platt, M. (1979). Propositions across utterances and speakers. In E. Ochs, & B. B. Schieffelin (Eds.), *Developmental Pragmatics* (pp. 251–268). Academic Press.

Office APAWP. (2006). Women in the American Psychological Association.

Office for National Statistics, UK Statistics Authority. (2021). Families by family type and presence of children. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/families/datasets/familiesandhouseholdsfamiliesandhouseholds

O'Meara, K., Kuvaeva, A., & Nyunt, G. (2017). Constrained choices: A view of campus service inequality from annual faculty reports. *The Journal of Higher Education*, *88*(5), 672–700.

Ota, M., Davies-Jenkins, N., & Skarabela, B. (2018). Why choo-choo is better than train: The role of register-specific words in early vocabulary growth. *Cognitive Science*, *42*(6), 1974–1999. https://doi.org/10.1111/cogs.12628

Pancsofar, N., & Vernon-Feagans, L. (2006). Mother and father language input to young children: Contributions to later language development. *Journal of Applied Developmental Psychology*, *27*(6), 571–587.

Pancsofar, N., Vernon-Feagans, L., & Investigators FLP. (2010). Fathers' early contributions to children's language development in families from low-income rural communities. *Early Childhood Research Quarterly*, *25*(4), 450–463.

Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, *13*(4), 539–545. https://doi.org/10.1016/0163-6383(90)90022-Z

Prolific Team. (2022). Prolific's payment principles.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). Look who's talking now! Parentese speech, social context, and language development across time. *Frontiers in Psychology*, *8*, 1008. https://doi.org/10.3389/fpsyg.2017.01008

Ranson, G. (2012). Men, paid employment and family responsibilities: Conceptualizing the 'working father'. *Gender, Work and Organization*, *19*(6), 741–761.

Roettger, T. B., Winter, B., Grawunder, S., Kirby, J., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in german. *Journal of Phonetics*, *43*, 11–25.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Schachter, F. F., Fosha, D., Stemp, S., Brotman, N., & Ganger, S. (1976). Everyday caretaker talk to toddlers vs. threes and fours. *Journal of Child Language*, *3*(2), 221–245. https://doi.org/10.1017/S030500090000146X

Seward, R. R., & Rush, M. M. A. (2015). Fathers, fathering, and fatherhood across cultures: Convergence or divergence?
Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (cog): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128.
Singh, L. (2018). He said, she said: Effects of bilingualism on cross-talker word recognition in infancy. *Journal of Child Language*, *45*(2), 498–510.
Siying, L., & Renji, S. (2019). Appreciating language conventions: Thirteen-month-old chinese infants understand that word generalization is shared practice. *Journal of Child Language*, *46*(4), 812–823.
Slaughter, V., & Suddendorf, T. (2007). Participant loss due to "fussiness" in infant visual paradigms: A review of the last 20 years. *Infant Behavior and Development*, *30*(3), 505–514.
Snow, C. E., & Ferguson, C. A. (1997). *Talking to children: Language input and acquisition*. Cambridge University Press. https://doi.org/10.2307/412603
Statistics Canada. (2014). Lone-parent families. https://www150.statcan.gc.ca/n1/pub/75-006-x/2015001/article/14202/parent-eng.htm
Statistics Canada. (2015). Same-sex couples and sexual orientation… by the numbers. https://www.statcan.gc.ca/en/dai/smr08/2015/smr08_203_2015
Stets, M., Stahl, D., & Reid, V. M. (2012). A meta-analysis investigating factors underlying attrition rates in infant erp studies. *Developmental Neuropsychology*, *37*(3), 226–252.
Stotzer, R. L., Herman, J. L., & Hasenbush, A. (2014). Transgender parenting: A review of existing research.
Sulpizio, S., Doi, H., Bornstein, M. H., Cui, J., Esposito, G., & Shinohara, K. (2018). fNIRS reveals enhanced brain activation to female (versus male) infant directed speech (relative to adult directed speech) in young human infants. *Infant Behavior and Development*, *52*, 89–96.
ter Haar, S. M., & Levelt, C. C. (2018). Disentangling attention for frequency and phonological markedness in 9-and 12-month-old infants. *Language Learning and Development*, *14*(4), 279–296.
Tornello, S. L., Riskind, R. G., & Babić, A. (2019). Transgender and gender non-binary parents' pathways to parenthood. *Psychology of Sexual Orientation and Gender Diversity*, *6*(2), 232–241. https://doi.org/10.1037/sgd0000323
U.S. Census Bureau. (2020). Families by presence of own children under 18: 1950 to present. census.gov/data/tables/time-series/demo/families/families.html
Van den Berg, W., Hendricks, L., Hatcher, A., Peacock, D., Godana, P., & Dworkin, S. (2013). 'One man can': Shifts in fatherhood beliefs and parenting practices following a gender-transformative programme in eastern cape, South Africa. *Gender and Development*, *21*(1), 111–125.
Van der Velde, B., & Junge, C. (2020). Limiting data loss in infant eeg: Putting hunches to the test. *Developmental Cognitive Neuroscience*, *45*, 100809.
Verdine, B. N., Bunger, A., Athanasopoulou, A., Golinkoff, R. M., & Hirsh-Pasek, K. (2017). Shape up: An eye-tracking study of preschoolers' shape name processing and spatial development. *Developmental Psychology*, *53*(10), 1869–1880.
Visser, I., Bergmann, C., Byers-Heinlein, K., Dal Ben, R., Duch, W., Forbes, S., Franchin, L., Frank, M. C., Geraci, A., Hamlin, J. K., Kaldy, Z., Kulke, L., Laverty, C., Lew-Williams, C., Mateu, V., Mayor, J., Moreau, D., Nomikou, I., Schuwerk, T., … Zettersten, M. (2022). Improving the generalizability of infant psychological research: The manybabies model. *Behavioral and Brain Sciences*, *45*, e35.
Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *43*(2), 230–246. https://doi.org/10.1037/h0084224
Winter, B., & Grice, M. (2021). Independence and generalizability in linguistics. *Linguistics*, *59*(5), 1251–1277. https://doi.org/10.1515/ling-2019-0049
Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, *45*, 1–37. https://doi.org/10.1017/S0140525X20001685

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Holtz, A., & Papineau, B. T. (2024). Examining the female-talker default in experimental language acquisition research. *Infant and Child Development*, *33*(1), e2475. https://doi.org/10.1002/icd.2475