Article

Taking Stock of the Evidence for the Gender-Equality Paradox in Gendered Names: A Reply to Berggren (2023) with New Data

Allon Vishkin[1] (iD)

Abstract

A phonetic feature called voicing has been shown to reflect the gendering of names. Vishkin et al. leveraged this insight to examine gender differentiation as a function of increasing gender equality, both across historical time and across the 50 United States. In this reply, I address a wide range of criticisms raised by Berggren on these findings. I begin by presenting novel data from 76 + million baby names in France from 1900 to 2021. Findings converge with Berggren's conclusion that the historical trend of voicing of female names is nonlinear and therefore cannot be fully accounted for by the monotonic increase of gender equality. However, I show the state-level analysis is robust to his critiques. I conclude that there are more gendered names in more gender-equal societies at the state level, even though the historical data does not shed light on the historical development of this phenomenon.

Keywords

gender, stereotypes, gender equality

A litany of grievances is raised by Berggren (2023) regarding the theoretical framework, analyses, and interpretation in Vishkin et al. (2022). To get to the root of the debate, I focus on critiques of the data and then address more general concerns beyond the data. In my response to some of the concerns, I include new data from more than 76 million baby names in France from 1900 to 2021. I conclude by summarizing the evidence regarding whether more gender-equal societies have more gendered names—briefly, the new French data, coupled with Berggren's re-analysis, leads me to conclude that the historical evidence regarding the role of gender equality in the gendering of names is weak. However, the cross-sectional evidence regarding the role of gender equality in the gendering of names is robust to Berggren's criticisms.

## Concerns About Study 1: United States Data

Is the Data Reliable?

Berggren argues that the data prior to 1937 is unreliable due to various potential selection biases. In particular, he suggests that the drop in voicing of female names from 80.4% in 1931 to 58.6% in 1970 is due to the exclusion of subsets of the population prior 1937. Given that the drop began several years before 1937 and continued for several decades after

1937, that account seems unlikely. Furthermore, Berggren suggests the drop is driven by the exclusion of certain groups prior to 1937, such as African Americans or domestic and agricultural workers. However, there is no clear reason why these subpopulations would have had substantially lower rates of voicing than the rest of the population. In addition, these subpopulations could not have accounted for the size drop because they were too small to account for the 21.8% drop (African Americans comprised approximately 9.5% of the population at that time [https://clinecenter.illinois.edu/project/Religious-Ethnic-Identity/composition-religious-and-ethnic-groups-creg-project]) or would have had to demonstrate massive differences in naming to drive the drop (Berggren suggests that workers excluded from the social security database accounted for 40% of the population, and for them to have shifted the percent of voiced female names from 80.4% in the rest of the population to 58.6% overall, they would have to had to have a voicing rate of female names of 25.9%).

[1]Technion—Israel Institute of Technology, Haifa, Israel

Corresponding Author:
Allon Vishkin, Faculty of Data and Decision Sciences, Technion—Israel Institute of Technology, Haifa 3200003, Israel.
Email: allonv@technion.ac.il
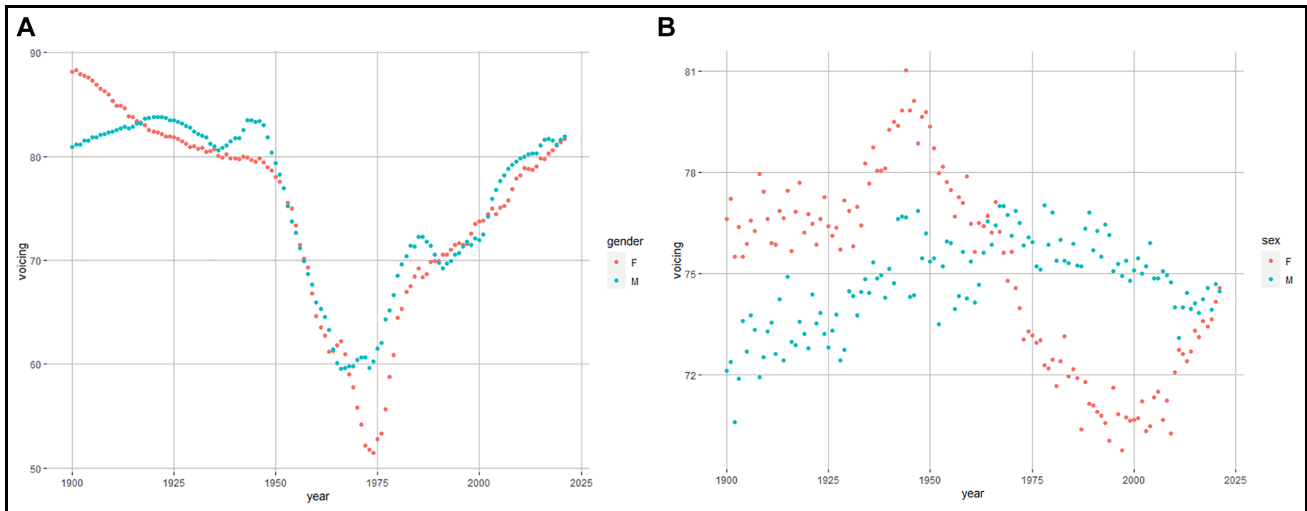
Figure 1. *The Percent of Voiced Names Given to Females and Males Across All Names (A) and for Unique Names (B)*

While it is not possible to directly compensate for selection biases in the U.S. data prior to 1937, the limitations of the historical U.S. findings can be assessed by testing whether a similar pattern of results occurs in a comparable data set. To do so, I analyzed data on the voicing of baby names from a novel data set of French names from 1900 to 2021. Data were downloaded from the French National Institute of Statistics and Economic Studies (www.insee.fr/fr/statistiques/2540004) on October 18, 2022. Data are based on the birth certificates of persons born in France from 1900 to 2021. After excluding names which have been masked by the curators of the data set (for being rare), names which are missing a birth year, or names from overseas departments (e.g., French Guiana in South America), the number of names included in the database is 76,778, 019. Exhaustiveness of the data is guaranteed by the curators only from 1946, coinciding with the end of World War II. Names were classified as voiced or voiceless based on their first phoneme as in Vishkin et al., with the exception that names beginning with *H* were classified as voiced, reflecting French pronunciation.

Figure 1A presents the percent of voiced names plotted for males and for females by year of birth. Since these results might be driven by a small set of highly popular names, Figure 1B presents the plot of the percent of voiced names when counting each unique name in each year once, so that every name receives equal weight regardless of the number of times it was given in a particular year. The pattern across the U.S. data (Figure 2 in Berggren, 2023) and the French data share several common features: (a) there is a drop in the voicing of female names in particular beginning in the middle of the 20th century, which continued for several decades; (b) this drop then reverses course, with female names become more voiced over the course of several decades; (c) both of these changes are more pronounced for females than for males; and (d)

changes in voicing tend to covary for both males and females.

The French data also suffer from its own potential selection biases prior to 1946. However, while the U.S. data are based on social security card applications, the French data are based on birth certificates. Furthermore, the beginning of exhaustive data in France coincides with the end of World War II. Finally, the significant drop in voicing of names in France begins only after exhaustiveness of the data is guaranteed, so potential selection biases in the French data cannot account for the main concern Berggren raised in the U.S. data. Thus, the convergence across the U.S. and French data, with each potentially suffering from a different set of selection biases, demonstrates that the results are very likely reliable above and beyond possible selection biases.

What is the Correct Analysis and Interpretation of the Data?

By plotting the data per gender and per year, Berggren correctly notes two problems with the analysis of the data in Vishkin et al. First, there is high auto-correlation of voicing across time which inflates Type-I errors. Second, the linearity assumption is not fulfilled, with an initial decrease in the voicing of female names, followed by the reverse trend. As such, linear regression is not an appropriate tool for analyzing the data. Furthermore, the data's nonlinear trend reveals a more complex association between historical time and the voicing of female names than a simple linear trend, indicating that any historical increases in gender equality—which has developed largely monotonically (Dorius & Firebaugh, 2010)—cannot fully explain the trend in voicing of female names over time in this data set. This is corroborated by the trend in the French data (Figure 1).

As Berggren notes, had the data points been plotted, then both the auto-correlation of datapoints and their non-linear trend would have been immediately apparent. How then, had I, as the sole author responsible for data analysis, missed this? My humbled response, which I offer by way of explanation rather than justification, is that I began analyzing these data as a third-year PhD student after having just made the transition from SPSS to R and was not versed at that time with data visualization—and even after having learned data visualization in R sometime later, I did not think to plot the data points because the effects were so large, and therefore—so it seemed—highly robust. I thus take sole responsibility for this error, and I am grateful to be living in the self-correcting era of open science.

Concerns About Study 1: English and Welsh Data

Does the English and Welsh Data in Study 1 Require Re-Analysis?

Berggren argues that the problem of auto-correlation is just as relevant to the analysis of the English and Welsh data in Study 1 as it is to the analysis of the United States data in Study 1. A Durbin–Watson test revealed that the residuals in the U.S. data are indeed correlated, for female names (auto-correlation = .977; Durbin–Watson statistic = .011; $p < .001$) and for male names (auto-correlation = .986; Durbin–Watson statistic = .025; $p < .001$). Meanwhile, residuals in the English and Welsh data are not correlated, neither for female names (auto-correlation = .018; Durbin–Watson statistic = 1.96; $p = .548$) nor for male names (auto-correlation = .004; Durbin–Watson statistic = 1.99; $p = .806$). Thus, the justification for re-analyzing and re-interpreting the U.S. data does not apply to the English and Welsh data, and the analyses and conclusions as originally reported in Vishkin et al. (2022) are thus valid.

Is There an Increase in the Voicing of Female Names Following 1970?

In addition to auto-correlation, Berggren attributes to the English and Welsh data the problem of a nonlinear trend. Specifically, Berggren's eyeballing of the data points leads him to conclude that female names became more voiced beginning around 1970 (p. 4 of the Supplemental Materials in Berggren, 2023). On average, female names do indeed become more voiced (1974: 57%; 1984: 59%; 1994: 66%). However, the differences are small, and Berggren conducts no significance tests. In fact, the difference in voiced female names is not statistically significant between 1974 and 1984, $\chi^2(1) = 0.08$, $p = .775$, between 1984 and 1994, $\chi^2(1) = 1.05$, $p = .307$, or even between 1974 and 1994, $\chi^2(1) = 1.71$, $p = .191$. Thus, contrary to Berggren's claim, there is no evidence that after the decline in the voicing of female names from 1904 to 1974, there is an increase in the voicing of female names.

Concerns About Study 2

Are the Simple Slopes Correct?

In Vishkin et al., the simple slope for males was reported to be significant while the simple slope for females was not significant, while Berggren finds that both slopes are significant and in the same direction (and the interaction remains significant). Re-running the analyses of simple slopes as they appear in the open-source script of Study 2 reveals a result which is consistent with Berggren's findings and inconsistent with those reported in Study 2. In particular, states with greater gender equality, as captured by leadership metrics, were more likely to give males a name beginning with a voiced phoneme, $b = .07$, $t = 5.54$, $p < .001$ (the original values reported in Vishkin et al: $b = .05$, $t = 2.48$, $p = .016$), and were also more likely to give females a name beginning with a voiced phoneme, $b = .04$, $t = 3.04$, $p = .003$ (the original values reported in Vishkin et al: $b = .01$, $t = 0.67$, $p = .50$), although the interaction remains significant and as originally reported. It is entirely unclear to me how the reporting error, for which I am solely responsible, crept into the text, particularly given their reproducibility from the study's open-access script. These corrected findings reveal that higher state-level gender quality predicts greater voicing of male names relative to female names (consistent with the conclusions of Vishkin et al), but that higher state-level gender equality nevertheless predicts greater voicing of female names as well (consistent with the findings of Berggren).

Do Gender Differences in Voicing Disappear When Controlling for States' Proportion of Foreign-Born Inhabitants?

Study 2 in Vishkin et al. showed that the proportion of voiced names for boys relative to girls is greater in more gender-equal states. Berggren argues that these results are driven by the higher proportion of foreign-born inhabitants in more gender-equal states, and provides support for this by controlling for the state-level proportion of foreign-born inhabitants as a covariate. However, these constructs suffer from high multi-collinearity and therefore are difficult to disentangle at the statistical level. A more direct test of Berggren's alternative account would involve excluding the most common names of foreign-born inhabitants from the data set. As described below, I conducted such a test and found no support for his alternative account.

Foreign-born residents of the United States originate from many different countries and speak various languages. Given that Mexican and other Latin-American countries made up about 50% of the regions of origin of foreign-born residents of the United States in 2018 (https://

www.pewresearch.org/hispanic/2020/08/20/facts-on-u-s-immigrants/), I re-ran the analyses when excluding the most common Hispanic names. I relied on a popular resource to identify 101 common names for girls and 100 common names for boys (www.babycenter.com/baby-names/most-popular/100-most-popular-hispanic-baby-names-of-2011_10363639; see Tables A1 and A2 in the Supplemental Materials). This led to the exclusion of 501, 786 names (17.1% of the sample). Re-running the analyses reported in Tables 1 and 2 of Vishkin et al. revealed highly similar results: seven of the eight interactions between gender and leadership equality are significant (see Tables A3 and A4 in the Supplemental Materials). Some effects are slightly smaller and others are slightly larger, but none are outside the confidence intervals of the original findings reported in Vishkin et al. These results are inconsistent with Berggren's argument that the greater gendering of names in more gender-equal countries is driven by the higher proportion of foreign-born inhabitants.

Why Focus on Gender Equality in Leadership?

Berggren questions why gender equality in leadership was considered the primary measure, compared to another measure of gender equality beyond leadership. Since this choice was not sufficiently justified in the target manuscript, I detail the implicit rationale below. There is a rich literature showing that achieving gender equality in leadership is a unique challenge (Heilman, 2001; Rudman & Glick, 1999, 2001) and that the contribution of women in leadership positions is also unique (e.g., Eagly & Johnson, 1990). Moreover, gender equality in leadership aggregates across indices of more visible types of equality (e.g., percentage of house or senate seats in state legislature held by women), while the measure of general gender equality includes indices which are much less visible (e.g., median pay ratio by gender, health care coverage and poverty level). The issue of visibility is critical to the mechanism proposed in section "General Discussion," the need to preserve distinctiveness, which should be activated to a greater extent in the context of highly visible gender equality.

Critiques Beyond the Data

Having addressed concerns about the reliability of the data and their interpretation, I turn to the larger concerns raised by Berggren beyond the data.

What Did Vishkin et al. Demonstrate?

The findings of Vishkin et al. are mischaracterized in the first sentence of Berggren's abstract. The findings did not demonstrate, or seek to demonstrate, "larger gender differences in voiced names [as a function of] higher gender equality". Instead, the findings demonstrated that greater gender equality predicts greater gendering of the names of males and females. The key distinction between Vishkin et al. and previous investigations of the gender-equality paradox is that Vishkin et al. operated under different assumptions: there are absolute criteria for the gendering of names based on voicing (more voicing = masculine; less voicing = feminine), and so the interesting question is not the divergence between voicing of names given to baby boys versus baby girls, but rather how that proportion of voicing has changed for each gender as gender equality has changed.

Thus, the key finding is not the relative difference between the voicing of male versus female names, but rather the trajectory of how gendered baby names are, as a function of gender equality. There is support for the motivational account of the gender-equality paradox if female names become less voiced at greater levels gender equality, as well as if male names have become more voiced. There is no reason to expect a symmetrical effect since female names follow different trends than male names (Varnum & Kitayama, 2011). The utility of testing for an interaction between Gender (names for baby girls vs. baby boys) × Time is not to show that the magnitude of the difference has been increasing, but rather to show that such a change over time is gender-specific. This is immediately apparent from Figure 1B in Vishkin et al., where gender differences are larger in 1904 than in 1994, but the trajectory across time shows greater gendering. Furthermore, Vishkin et al. state this explicitly in the predictions: "We predicted that voiced names would be increasingly given to males and unvoiced names would be increasingly given to females over time" (p. 491).

Do Previous Findings Fail to Show a Gender-Equality Paradox Over Time?

As Berggren indicates, it is important to place research findings in the context of the larger literature. Is there any evidence that gender differences have increased over time? Berggren cites a single source as providing the only relevant evidence to bear on this question, and that source found "a decreased differentiation with time". In fact, several studies have found an *increase* in gender differentiation with time, including in adolescent mental health (Högberg et al., 2020; Thorisdottir et al., 2017), happiness (Stevenson & Wolfers, 2009), and cognitive abilities (Weber et al., 2014).

Summarizing the Evidence: Do More Gender-Equal Societies Have More Gendered Names?

The French historical data from 1900 to 2021 converges with Berggren's presentation of the historical U.S. data from 1880 to 2018 by showing a drop in the voicing of female names in particular beginning in middle of the 20th century, which continued for several decades, followed by a

course reversal, with female names becoming more voiced over the course of several decades. Given that gender equality has increased largely monotonically over time (Dorius & Firebaugh, 2010), these nonlinear changes cannot be fully accounted for by gender equality, if they are accounted for by gender equality at all. The historical English and Welsh data from 1904 to 1994 shows only a drop in the voicing of female names, with no increase, but the inability to detect an increase might be due to the smaller data set in that sample, with data available only every 10 years. Thus, the three historical datasets point to weak evidence at best regarding the role of gender equality in the gendering of names.

In contrast to the historical datasets, the cross-sectional finding that the voicing of male names is more gendered in more gender-equal states, relative to the voicing of female names, has withstood several robustness checks, including in Vishkin et al. and in criticisms raised by Berggren. In Vishkin et al., these findings held with and without several controls, including state-level differences in statehood, state-level sex ratios, and state-level population. They held across all names as well as unique names, and replicated in data for 2017, for 2018, and for 2019. Moreover, contrary to Berggren's alternative account that these results are driven by the presence of foreign-born inhabitants in more gender-equal states, these findings held when excluding common Hispanic names.

In conclusion, the cross-sectional state-level data reveals that there are more gendered names in more gender-equal societies, even though the historical data does not shed light on the historical development of this phenomenon.

Author's Note

Scripts are available at https://osf.io/587ep/.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Allon Vishkin https://orcid.org/0000-0002-9655-7449

Supplemental Material

Supplemental material for this article is available online.

References

Berggren, M. (2023). No evidence of a gender-equality paradox in gendered names: Comment on Vishkin, Slepian, and Galinsky (2022). *Social Psychological and Personality Science*. Advance online publication. https://doi.org/10.1177/1948550622113 4353

Dorius, S. F., & Firebaugh, G. (2010). Trends in global gender inequality. *Social Forces*, *88*(5), 1941–1968. https://doi.org/10.1353/sof.2010.0040

Eagly, A. H., & Johnson, B. T. (1990). Gender and leadership style: A meta-analysis. *Psychological Bulletin*, *108*(2), 233–256.

Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*(4), 657–674. https://doi.org/10.1111/0022-4537.00234

Högberg, B., Strandh, M., & Hagquist, C. (2020). Gender and secular trends in adolescent mental health over 24 years—The role of school-related stress. *Social Science and Medicine*, *250*, Article 112890. https://doi.org/10.1016/j.socscimed.2020.112890

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*(5), 1004–1010. https://doi.org/10.1037/0022-3514.77.5.1004

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*(4), 743–762. https://doi.org/10.1111/0022-4537.00239

Stevenson, B., & Wolfers, J. (2009). The paradox of declining female happiness. *American Economic Journal: Economic Policy*, *1*(2), 190–225. https://doi.org/10.1257/pol.1.2.190

Thorisdottir, I. E., Asgeirsdottir, B. B., Sigurvinsdottir, R., Allegrante, J. P., & Sigfusdottir, I. D. (2017). The increase in symptoms of anxiety and depressed mood among Icelandic adolescents: Time trend between 2006 and 2016. *European Journal of Public Health*, *27*(5), 856–861. https://doi.org/10.1093/eurpub/ckx111

Varnum, M. E. W., & Kitayama, S. (2011). What's in a name? Popular names are less common on frontiers. *Psychological Science*, *22*, 176–183.

Vishkin, A., Slepian, M. L., & Galinsky, A. D. (2022). The gender-equality paradox and optimal distinctiveness: More gender-equal societies have more gendered names. *Social Psychological and Personality Science*, *12*(2), 490–499.

Weber, D., Skirbekk, V., Freund, I., & Herlitz, A. (2014). The changing face of cognitive gender differences in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(32), 11673–11678. https://doi.org/10.1073/pnas.1319538111

Author Biography

Allon Vishkin is an assistant professor at the Technion – Israel institute of Technology. He studies cultural differences in emotion and motivation.

Handling Editor: Jennifer Bosson