

Titus Raban von der Malsburg | Till Poppels | Roger P. Levy

## **Implicit gender bias in linguistic descriptions for expected events: the cases of the 2016 United States and 2017 United Kingdom elections**

**Suggested citation referring to the original publication:**

**Psychological Science 31 (2020) 2, pp. 115 - 128**

**DOI: <https://doi.org/10.1177/0956797619890619>**

**ISSN: 0956-7976**

**Journal article | Version of record**

**Secondary publication archived on the Publication Server of the University of Potsdam:**

**Zweitveröffentlichungen der Universität Potsdam : Humanwissenschaftliche Reihe 856**

**ISSN: 1866-8364**

**URN: <https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-516154>**

**DOI: <https://doi.org/10.25932/publishup-51615>**

**Terms of use:**

**This work is licensed under a Creative Commons License. This does not apply to quoted content from other authors. To view a copy of this license visit**

**<https://creativecommons.org/licenses/by/4.0/>.**



# Implicit Gender Bias in Linguistic Descriptions for Expected Events: The Cases of the 2016 United States and 2017 United Kingdom Elections



Titus von der Malsburg<sup>1,2</sup> , Till Poppels<sup>3</sup>, and Roger P. Levy<sup>2,3</sup>

<sup>1</sup>Department of Linguistics, University of Potsdam; <sup>2</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology; and <sup>3</sup>Department of Linguistics, University of California San Diego

Psychological Science  
2020, Vol. 31(2) 115–128  
© The Author(s) 2020



Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0956797619890619  
www.psychologicalscience.org/PS



## Abstract

Gender stereotypes influence subjective beliefs about the world, and this is reflected in our use of language. But do gender biases in language transparently reflect subjective beliefs? Or is the process of translating thought to language itself biased? During the 2016 United States ( $N = 24,863$ ) and 2017 United Kingdom ( $N = 2,609$ ) electoral campaigns, we compared participants' beliefs about the gender of the next head of government with their use and interpretation of pronouns referring to the next head of government. In the United States, even when the female candidate was expected to win, *she* pronouns were rarely produced and induced substantial comprehension disruption. In the United Kingdom, where the incumbent female candidate was heavily favored, *she* pronouns were preferred in production but yielded no comprehension advantage. These and other findings suggest that the language system itself is a source of implicit biases above and beyond previously known biases, such as those measured by the Implicit Association Test.

## Keywords

language, psycholinguistics, event expectations, reference, implicit bias, open data, open materials

Received 9/27/18; Revision accepted 9/26/19

When reading the sentence, “The nurse prepared himself for the operation,” you may find yourself experiencing a brief hint of surprise at the word “himself,” even though you know that many nurses are male. This surprise is widely experienced and amenable to experimental measurement, such as word-by-word reading times (Duffy & Keir, 2004; Foertsch & Gernsbacher, 1997; Kennison & Trofe, 2003; Sturt, 2003) and electroencephalographic response (Osterhout, Bersick, & McLaughlin, 1997). In the United States (U.S.), the United Kingdom (U.K.), and many other societies, there is a *stereotype*—a belief about the typical characteristics of a group (Judd & Park, 1993)—that nurses are female; this stereotype rapidly influences readers' expectations about how the sentence may unfold, including the gender of a pronoun referring to the nurse. These stereotypes are reflected in the statistics of texts (Caliskan, Bryson, & Narayanan, 2017), and like many other factors (Altmann

& Steedman, 1988; Bicknell, Elman, Hare, McRae, & Kutas, 2010; Kuperberg & Jaeger, 2016; Marslen-Wilson, 1975; Nieuwland & Van Berkum, 2006; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Traxler, 2014) affect real-time language comprehension and production (collectively, *language processing*) rapidly and without requiring introspection or awareness—an implicit biasing effect (Greenwald & Banaji, 1995; Greenwald & Krieger, 2006; Greenwald, McGhee, & Schwartz, 1998).

Despite numerous studies documenting gender-bias effects in language processing, our understanding of these effects remains limited. The biasing influence of

---

## Corresponding Author:

Titus von der Malsburg, University of Potsdam, Department of Linguistics, Karl-Liebknecht-Strasse 24-25, 14476 Potsdam, Germany  
E-mail: malsburg@uni-potsdam.de

gender stereotypes on language processing might be entirely mediated by expectations about events in the world. For example, in the U.S. in 2018, more than five out of six nurses were female (Bureau of Labor Statistics, 2018), and more generally, gender stereotypes are well calibrated to true gender ratios (Garnham, Doehren, & Gygax, 2015). Reading about a nurse without further preceding context might most likely lead one to imagine the nurse as a woman and expect any pronouns referring to the nurse to be *she* pronouns, leading to surprise at the pronoun “himself.” We might hypothesize that event expectations fully mediate the biasing effect of stereotypes on linguistic preferences in this way, with linguistic preferences transparently reflecting event expectations. In this case, the mapping between event expectations and linguistic preferences in pronoun usage should be unbiased (and thus symmetric) with respect to gender. This hypothesis makes a testable prediction. For prominent events that involve specific, well-known people and command widespread attention in an experimental population, expectations regarding those events may be directly measurable. If the hypothesis is true, we should find that linguistic preferences regarding descriptions of those events are unbiased relative to these measured event expectations: Equal-strength expectations for a referent to be a given gender will lead to equal strengths of preference for matching-gender pronouns, regardless of whether that gender is female or male. Alternatively, we might hypothesize that the mapping between event expectations and linguistic preferences itself is biased: Equal-strength expectations for referent gender might, for example, yield a stronger preference for *he* pronouns when that gender is male than for *she* pronouns when that gender is female.

We took advantage of the 2016 U.S. presidential campaign and the 2017 U.K. general election to investigate this mapping experimentally. Electoral campaigns for a national head of government attract tremendous interest, attention, and news coverage, leading to expectations on the part of citizens regarding likely outcomes that fluctuate with campaign events and news coverage. In the U.S., the 2016 presidential campaign presented the concrete prospect of a woman, Hillary Clinton, as the next head of state and government for the first time in the country’s history. In the 2017 U.K. general election campaign, the favored incumbent head of government was a woman, Theresa May. In both cases, the competing candidates were male. These cases offered natural experiments allowing us to cleanly investigate how closely preferences regarding pronominal references to the next head of government match beliefs about who would win the election.

We collected experimental data in three tasks: *belief estimation* to determine whom the participant expected to win the election, *text completion* to measure preferences in language production regarding which pronoun to use to refer to the next head of government, and *self-paced reading* to measure (using reading times) the relative ease of comprehension of pronominal references to the next head of government. Belief estimation yields probability estimates of expectations regarding whether the next head of state will be female or male; the other two tasks reveal pronoun preferences in language production and comprehension respectively, which can be compared with these event expectations.

If event expectations fully mediate the influences of gender biases, and linguistic preferences transparently reflect those expectations, the relationship between event expectations and linguistic preferences should look the same from the standpoint of either female or male gender. In production, for example, if 40% of pronouns are *he* when the male candidate is estimated to have a 55% chance of winning, then 40% of pronouns should be *she* when the female candidate in the same election is estimated to have a 55% chance of winning (Fig. 1a; note that the gendered-pronoun rate may be lower than the corresponding event-expectation rate because other pronouns, such as singular *they*, may also be used). In comprehension, pronouns matching the gender of the candidate expected to win should be less surprising and thus less disruptive to encounter than gender-mismatching pronouns.

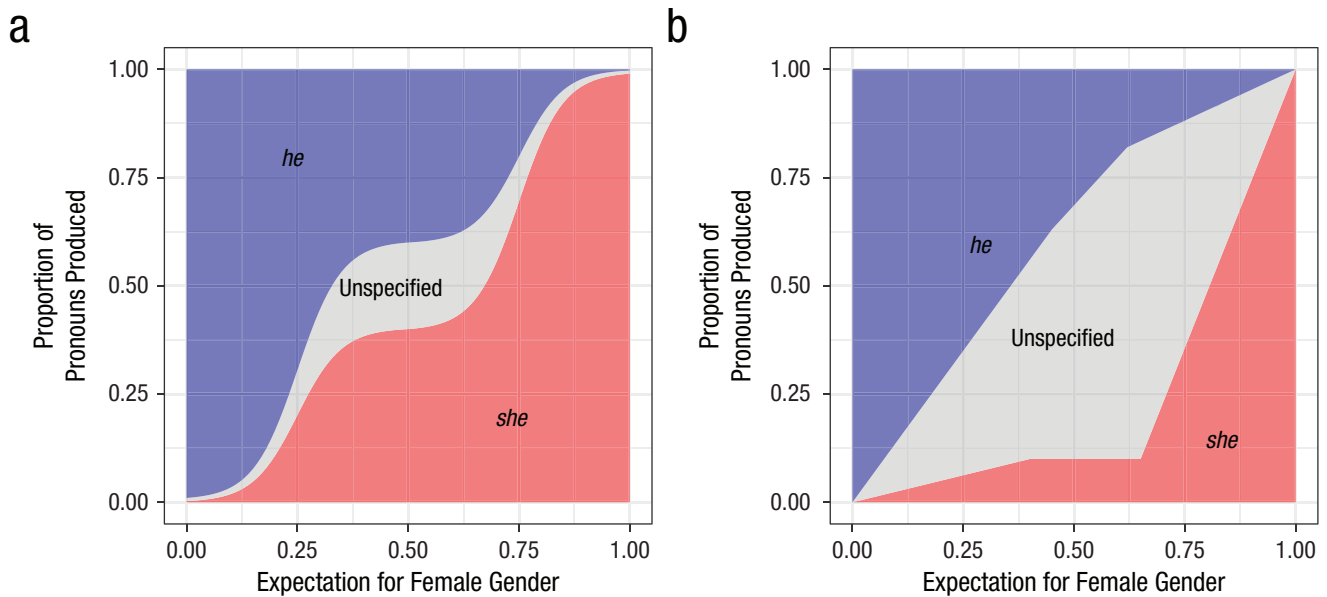
If, in contrast, implicit gender bias affects not only expectations about world events (Greenwald et al., 1998) but also how these expectations are translated into linguistic forms, then the relationship between event expectations and linguistic preferences may differ depending on the gender considered. For example, *she* pronouns might be disfavored relative to event expectations because of historical and contemporary social gender stereotypes, both those particular to the role of head of government and more broadly (Fig. 1b).

For both campaigns, we collected data across multiple experimental rounds, allowing us to see whether and how quickly changes in beliefs are reflected in linguistic preferences.

## Experiment 1: The 2016 U.S. Presidential Election Campaign

### Method

Between June 2016 and January 2017, we collected 12 rounds of data relating to the U.S. presidential campaign using three single-trial tasks: belief estimation,



**Fig. 1.** An example unbiased mapping (a) and an example biased mapping (b) of gender expectations to the proportion of pronouns produced. In (a), the mapping from female-gender event expectations to *she* pronoun probabilities is the same as that from male-gender expectations to *he* pronoun probabilities. Because this mapping is unbiased in this sense, it is symmetric under 180° rotation. In (b), a biased mapping consistent with population-average behavior in Experiment 1 (the United States election) is shown. This mapping is not symmetric under 180° rotation. In both examples, “unspecified” refers to all pronominal forms that are gender nonspecific, including *they*, *he* or *she*, *he/she*, and so forth.

text completion, and self-paced reading. In the belief-estimation task, participants reported their belief in the likelihood that each of the three candidates who were still plausibly in contention at the start of our study—Hillary Clinton, Bernie Sanders, and Donald Trump—would win the election. In the text-completion task, participants completed a two-sentence text fragment designed to elicit a pronominal reference to the next U.S. president. In the self-paced reading task, participants read a three-sentence vignette. The first sentence was a stage-setting sentence, and the second and third sentences included pronominal references to the future president. Participants read these texts using a self-paced moving window (Mitchell, 1984), allowing us to measure word-by-word processing times and, specifically, the reading times elicited by pronouns.

**Participants.** Self-reported U.S. citizens ( $N = 24,863$ ) from a broad population were recruited via Amazon Mechanical Turk (Levay, Freese, & Druckman, 2016; Stewart, Chandler, & Paolacci, 2017). All participants gave informed consent in accordance with the regulations of the University of California San Diego Institutional Review Board. Compensation was calibrated to yield approximately federal minimum wage (\$7.25 per hour). To avoid cross-trial and cross-task influences, we designed the experiment so that every participant participated only once and only in one of the three tasks. We further solicited demographic

information and frequency of election-related news consumption (see Table S2 in the Supplemental Material available online). In every round, more than 70% of participants reported daily news consumption, and at most 5% reported less than weekly consumption.

**Apparatus.** The belief-estimation task was implemented using Qualtrics, an online survey platform (<https://www.qualtrics.com/>). The browser-based software Ibex was used to implement the text-completion task and the self-paced reading task (<https://github.com/addrummond/ibex>).

**Materials.** Stimuli in the text-completion task and the self-paced reading task started with a fixed stage-setting sentence: “The next US president will be sworn into office in January 2017.” This sentence introduced the topic—the future president of the U.S.—and made it clear that the president in question was the winner of the ongoing race.

Twelve target sentences were designed for the production and comprehension tasks. These sentences described activities that are highly stereotypical for U.S. presidents and each included one pronoun referring to the next U.S. president. There were no other nouns in these sentences that the pronouns could refer to. See the Supplemental Material for the full list of sentences.

In the text-completion task, the stage-setting sentence was followed by a sentence fragment that ended

just before where a pronoun would be expected (e.g., “The next US president will be sworn into office in January 2017. After moving into the Oval Office, one of the first things that . . .”). The target sentences were designed so that readers would be likely to continue these fragments with a pronoun reference to the next president when completing the text (and to expect a pronoun reference to the next president during comprehension).

In the self-paced reading task, the full target sentences were used with *she*, *he*, or singular *they*. In addition to measuring ease of comprehension when a pronoun reference to the next president was first encountered, we were interested in the effect of the pronoun on subsequent comprehension behavior. Would establishing the gender of the next president through an explicit *she* or *he* reference eliminate any influence of event expectations or biases? To address this question, we combined each of the 12 pronoun sentences with each of the remaining 11 pronoun sentences to form 132 sentence triplets (including the stage-setting sentence). Five experimental conditions were formed. The first four conditions crossed *she* and *he* in the second and third sentence such that we had a pair of pronoun-matching conditions (*he . . . he*, *she . . . she*) and another pair of pronoun-mismatching conditions (*she . . . he*, *he . . . she*). The fifth condition used the gender-neutral singular *they* in both pronoun sentences (*they . . . they*). The following is an example sentence from the *he . . . she* condition of the task:

The next US president will be sworn into office in January 2017. After moving into the Oval Office, one of the first things that he will do is hold a staff briefing. During the inauguration speech, the president will emphasize her commitment to resolve outstanding issues quickly.

If gender information from the first pronoun completely overrides event expectations and any broader biases, reading times for the third sentence should be determined solely by whether the gender of the pronoun in that sentence matches the gender of the pronoun in the second sentence. Effects beyond gender match to the previous pronoun might provide further evidence that comprehenders’ knowledge and beliefs do not transparently map to linguistic preferences; rather, linguistic preferences might themselves be subject to additional biases.

A secondary purpose of the pronoun-match versus pronoun-mismatch manipulation was to provide a benchmark for evaluating the size of gender-bias effects for the first pronoun, because gender mismatches within a text (e.g., when a person is first referred to as

*she* and then as *he*) are expected to elicit severe processing difficulty.

After the election and Donald Trump’s inauguration, minor changes were introduced into the sentences to ensure consistency with real-world events that had transpired. This affected mostly tense (e.g., “After the inauguration, [he/she/they] will do” became “. . . [he/she/they] did”). No such changes were necessary during the preelection period.

**Procedure.** Participants started the experiment by opening a link in a new browser window. After participants gave informed consent, they were asked how closely they were following the news about the presidential race (multiple-choice options were *daily*, *weekly*, *monthly*, *less than monthly*, *never*). Then the experimental task started.

Each participant performed only one of the three tasks (belief estimation, text completion, self-paced reading). Further, in the text-completion task and in the self-paced reading task, they performed only one trial (i.e., they were presented with only 1 of the 12 sentence fragments for completion or only 1 of the 132 texts for reading).

In the belief-estimation task, participants reported how likely they thought it was that each of the three candidates—Hillary Clinton, Bernie Sanders, and Donald Trump—would win the election by indicating a response between 0 and 100 using sliders. We included Sanders at the start of the study because Clinton had not yet secured sufficient delegates to ensure the Democratic Party presidential nomination, and we kept Sanders in during subsequent rounds for consistency. Participants were allowed to position each slider freely. The order of candidates was randomized. We converted each participant’s normalized response to a degree of belief ranging from 0 to 100 that the next president would be female by dividing the response given for Clinton by the sum of the responses for all three candidates.

We should note that although normalized slider scores are a preferred method for this type of belief-estimation task (Franke et al., 2016), evidence suggests that participants are often reluctant to give extreme slider scores. For example, the mean normalized belief that Sanders would be the next president hovered around 5% even after the Democratic convention, when Sanders was no longer in the race and had endorsed Clinton (see Fig. S3 in the Supplemental Material). As a result, our procedure is likely, if anything, to underestimate the strength of the belief in our experimental population that the next president would be female. Comparison with historical betting markets suggests that this conclusion is plausible (Real Clear Politics, 2016). If we underestimated this belief, it would imply

that the degree of bias against *she* pronoun usage relative to beliefs about world-event likelihoods was even stronger than our data suggest at face value.

In the text-completion task, participants were presented with the stage-setting sentence and 1 of the 12 sentence fragments. Completions could be free form and arbitrarily long.

In the self-paced reading task, participants first saw horizontal lines that masked the complete text, including the white spaces between words. Words had to be revealed one by one by pressing the space bar. Each time a word was revealed, the previous word was masked again, so just one word was visible at a time. Before the experimental sentences were presented, participants read two practice sentences. After participants finished reading the experimental sentences, they were presented with two yes/no comprehension questions—one for each pronoun sentence—whose purpose was to confirm that they paid attention to the text. Next, participants were asked to rate the “naturalness of the text passage” on a scale from 1 (*makes no sense at all*) to 7 (*makes perfect sense*).

After every task, participants completed a demographics survey. Finally, participants were debriefed and received a participation code that they could use to confirm completion of the study on the Mechanical Turk website. On average, participants completed the experiment in approximately 3 min.

**Data collection.** We collected 10 preelection rounds of data, beginning shortly before the June 7 California Democratic-party primaries and ending shortly before the October 25 third presidential debate (Fig. 2). Each round began on a Sunday morning. Nine of these rounds were collected at predetermined times preceding, following, or both preceding and following significant milestones in the race, such as the party conventions and the TV debates. We added an ad hoc round of data collection during the aftermath of Donald Trump’s confrontation with the Muslim family of the killed U.S. soldier Humayun Khan. We also collected two rounds of postelection data, one shortly after the November 8 election and one shortly after the January 20 inauguration. We collected no pre-election data after the third debate in order to avoid any possible depletion of the participant pool for immediate postelection data collection. Conservative estimates of the number of U.S.-based workers on Mechanical Turk suggested that we might be exhausting the pool when the election was approaching. Consistent with this idea, the data showed that there was a slight slowdown in participation in the ninth round (before the third TV debate).

The amount of data collected at each of these times was determined on the basis of two pilot experiments carried out in April and May 2016. These pilot experiments were similar in design to the main experiment

but had smaller numbers of stimuli and participants. In each data-collection round of the main experiment, we aimed for 280 participants in the belief-estimation task, 560 participants in the text-completion task, and 1,120 participants in the self-paced reading task. Because of how tasks are advertised on Mechanical Turk, the precise numbers of tested participants varied slightly around these target values.

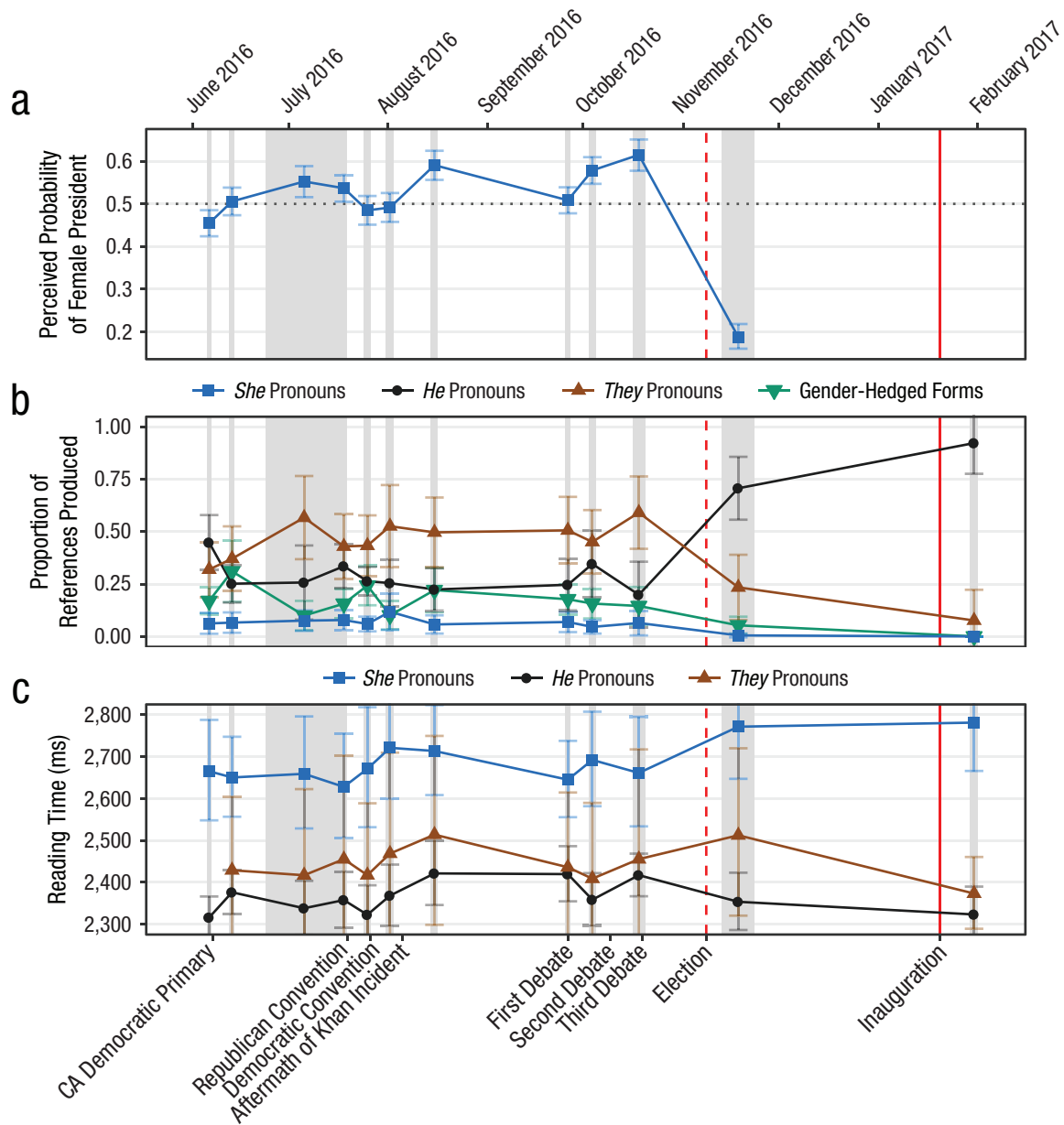
We took multiple precautions to avoid any possible differences in the populations tested across the three tasks. To avoid differences arising from the time of day and the day of the week, we tested participants in all three tasks concurrently. Specially designed software was used to make sure the testing periods were the same across tasks despite the differing numbers of participants per task. To avoid population differences from self-selection biases, we advertised all three tasks in the exact same way on the crowdsourcing platform. To avoid population differences arising from differences in monetary compensation, we paid participants the same amount for completing all three tasks (\$0.35).

**Data preprocessing.** We removed participants who indicated that they were noncitizens of the U.S., nonresidents of the U.S., or nonnative speakers of English (~7%). We also removed some participants who participated multiple times during a brief period when our mechanism for preventing multiple participation temporarily failed.

The scores assigned to each candidate in the belief-estimation task were converted to proportions, and the proportion assigned to Clinton was interpreted as the participant’s implicit belief that a woman would become president.

In the text-completion task, pronouns referring to the next president were automatically detected using a regular-expression search. Pronouns were used in 41% of the sentence completions. Among the remaining completions, the most common were those without animate noun phrases (33%; e.g., “During the first term, the president will have to work towards . . . immigration reform”) and full noun-phrase references compatible with any future president (11%; e.g., “the president of the United States”). Pronouns were categorized into the following four types: *she* (the feminine forms *she/her*), *he* (the masculine forms *he/his/him*), *they* (the forms *they/their/them*), and *gender-hedged* (*he or she*, *he/she*, and variants).

From the data collected with the self-paced reading task, we removed trials on which participants had implausibly short (< 180 ms) or long (> 5,000 ms) reading times on any of the words. This affected 19% of the trials. Per-word reading times were residualized to remove known effects of factors such as word length, word position in the text, and punctuation following a



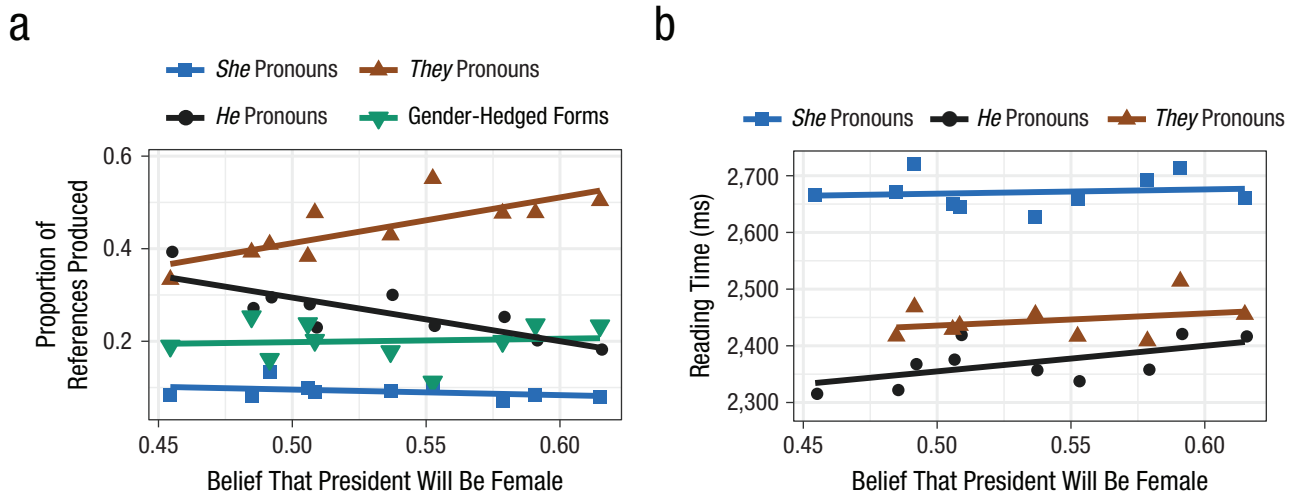
**Fig. 2.** Results from Experiment 1 from before to after the election: (a) implicit beliefs that the next president would be female, (b) proportion of each type of pronoun produced to refer to the next president, and (c) summed reading time (geometric mean) for the first pronoun and the four subsequent words, separately for each type of pronoun read. Gender-hedged forms include “he or she,” “he/she,” and similar variants. Error bars show 95% confidence intervals. In (a), the dotted horizontal line indicates equally balanced expectations that the next president would be female or male. In all graphs, the gray bands indicate data-collection periods, and the red lines indicate the dates of the election (dashed line) and inauguration (solid line).

word (for full details, see the description of Experiment 0 in the Supplemental Material). All reading times reported in this article are residualized.

**Data analysis.** Analyses were conducted using Bayesian hierarchical generalized linear models with maximal random-effects structure (Barr, Levy, Scheepers, & Tily, 2013) implemented in the Stan system for full Bayesian inference (Carpenter et al., 2017). We report the posterior

mean of each parameter in question, the 95% credible interval (CrI), and the posterior probability of the parameter being on the same side of zero as its posterior mean. We log-transformed self-paced reading times before analysis, but we also report estimated effect sizes in milliseconds, which were calculated by taking the differences between estimated condition means after back-converting them to milliseconds. See the Supplemental Material for more details.





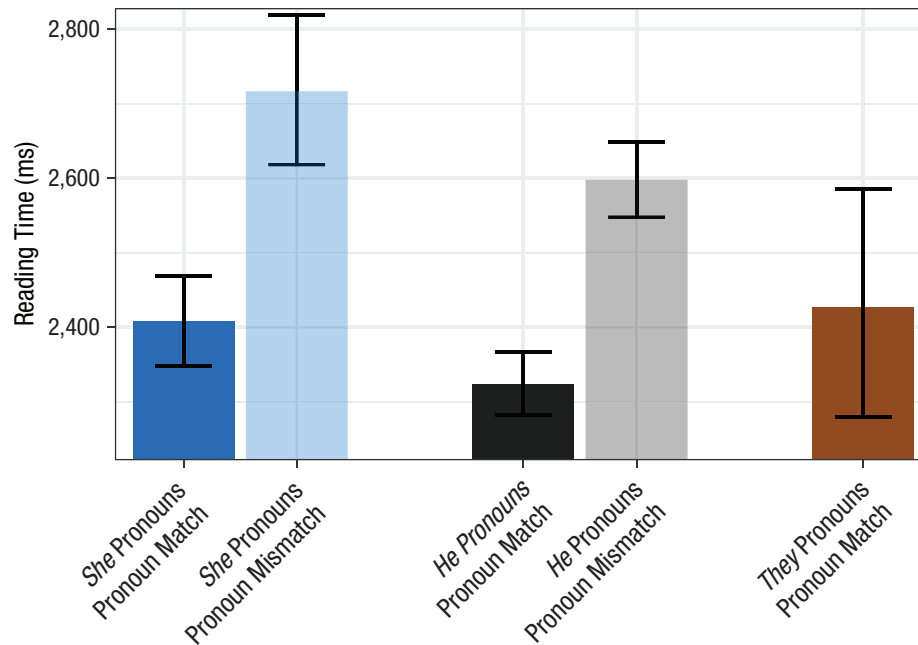
**Fig. 3.** Results from Experiment 1: pronoun processing as a function of implicit beliefs about the future president's gender. The proportion of pronoun references produced (a) is shown for each of the four pronoun types. Summed reading time for the first pronoun and the four subsequent words (b) is shown for each of the three pronoun types used in the stimulus sentences. Each data point represents a by-condition mean for one of the 10 rounds of data collected before the election. Gender-hedged forms include "he or she," "he/she," and similar variants. Slopes indicate best-fitting regressions.

## Results

Participant demographics indicated that the pool of participants was diverse but not entirely representative of the general population of the U.S. For instance, the median age of participants in our study was 32 years, compared with 38 years in the general U.S. population (He, Goodkind, & Kowal, 2016). Also, 57% of our participants identified as female, whereas women constitute only 51% of the U.S. population (Howden & Meyer, 2011). There were some small changes in demographics across the data-collection period, but these were within the range of what might be expected on the basis of random fluctuations (Casey, Chandler, Levine, Proctor, & Strolovitch, 2017). See the Supplemental Material for more complete details on participant demographics.

The average participant expectation for a female president was significantly above 50% over much of the preelection period and continued to grow as the election approached (Fig. 2a),  $\hat{\beta} = 0.082$ , 95% CrI =  $[-0.0037, 0.17]$ ,  $P(\beta > 0) = .97$  in a beta regression with  $\hat{\beta}$  being the estimated log-odds increase per month. Note that expectations for a female president in the immediate postelection experimental round remained at 17%. Such a high expectation could reflect limitations of our estimation method, but it could also reflect the surprise and disbelief at the reported election results immediately following the election (some election forecasting sites had reported the probability of a Clinton win as well above 90%, Wang, 2016, and recount efforts were soon to begin in several swing states).

Text-completion results are shown in Figure 2b as a function of time and Figure 3a as a function of the strength of the expectation for a female president. Before the election, *they* was the most common pronoun type for all but the first experimental round. Strikingly, participants produced *she* far less often than *he*, even when expectations favored a female president. Production of *he* was negatively correlated with the expectation of a female president across the preelection experimental rounds,  $\hat{\beta} = -1.4$ , 95% CrI =  $[-2.9, -0.047]$ ,  $P(\beta < 0) = .98$  in a logistic regression with  $\hat{\beta}$  being the estimated log-odds increase for one unit of the predictor (i.e., expectation). However, as expectations for a female president strengthened, participants gravitated not toward *she*,  $\hat{\beta} = -0.25$ , 95% CrI =  $[-1.6, 1.2]$ ,  $P(\beta < 0) = .66$ , but toward *they*,  $\hat{\beta} = 1.2$ , 95% CrI =  $[0.28, 2.2]$ ,  $P(\beta > 0) > .99$ . These patterns held regardless of participant age, gender, education, and political-party affiliation (see Figs. S20–S29 in the Supplemental Material), and they demonstrate a biased mapping from election-outcome expectation to preferred linguistic form: Relative to referent-gender expectations, *she* production rates were lower than *he* production rates. Although the *she* disadvantage could in part reflect a default (and perhaps arguably unbiased) preference among some speakers for *he* whenever the referent's gender is uncertain (generic *he*), closer analysis indicates that such a default preference cannot plausibly provide a complete explanation for our results (see "Incompatibility of Production Data With an Unbiased Expectation-to-Language Mapping" in the Supplemental Material).<sup>1</sup>



**Fig. 4.** Results from Experiment 1: summed reading time for the second pronoun and the four subsequent words, separately for each pronoun type and pronoun-match condition. Results shown are from only the 10 rounds of data collected before the election. Means are geometric, and error bars indicate 95% confidence intervals.

Reading times in response to the first pronoun offer insight into how subjective event expectations are translated to language-comprehension preferences, closely analogous with text completion. Additionally, reading times in response to the second pronoun offer insight into the role of a third information source—referent gender as established by the first pronoun. We analyzed reading times, summed over each pronoun and its four succeeding words, to capture the spillover of processing difficulty often observed in self-paced reading experiments (Mitchell, 2004; Smith & Levy, 2013). See Figures S4 and S7 in the Supplemental Material for word-by-word results.

Results for the first pronoun in the self-paced reading task are shown in Figure 2c as a function of experimental round and in Figure 3b as a function of the strength of the expectation of a female president. *She* was read significantly slower than *he*,  $\hat{\beta} = 0.13$ ,  $\Delta$ reading time = 302.31 ms, 95% CrI = [0.09, 0.17],  $P(\beta > 0) > .99$  in a linear regression of log-reading times with  $\hat{\beta}$  being the approximate estimated proportional difference (i.e., 13% in this case), and *she* was also read significantly slower than *they*,  $\hat{\beta} = 0.10$ ,  $\Delta$ reading time = 250.90 ms, 95% CrI = [0.01, 0.18],  $P(\beta > 0) = .99$ , across all experimental rounds. Reading times for *he* and *they* did not significantly differ from each other before the election,  $\hat{\beta} = 0.03$ ,  $\Delta$ reading time = 70.40 ms, 95% CrI = [-0.06, 0.12],

$P(\beta > 0) = .77$ . Across preelection experimental rounds, *he* reading times were positively correlated with expectation for a female president,  $\hat{\beta} = 0.047$ , 95% CrI = [-0.0097, 0.11],  $P(\beta > 0) = .95$  (Fig. 3b), but neither *she* nor *they* showed any effect of changes in expectation,  $\hat{\beta} = 0.013$ , 95% CrI = [-0.05, 0.069],  $P(\beta > 0) = .67$ ;  $\hat{\beta} = 0.032$ , 95% CrI = [-0.058, 0.12],  $P(\beta > 0) = .77$ . As in the text-completion task, these patterns were stable across the demographic spectrum (see Figs. S20–S29).

Second pronouns were read much slower when they mismatched the gender of the first pronoun than when they matched, indicating a strong effect of gender information provided by the preceding text (Fig. 4),  $\hat{\beta} = 0.11$ ,  $\Delta$ reading time = 282.13 ms, 95% CrI = [0.09, 0.14],  $P(\beta > 0) > 0.99$ . Note that the processing disadvantage for *she* first pronouns was of similar size: 302 ms for a first pronoun mismatching the gender stereotype versus 282 ms for a second pronoun mismatching the gender information provided by the first pronoun. This suggests that, in terms of processing cost, the disruption from an initial *she* was just as strong as the disruption from inconsistency of pronoun gender for the same referent from sentence to sentence.

Furthermore, we saw a sizeable processing disadvantage when the second pronoun was *she*. *She* was read slower than *he* overall,  $\hat{\beta} = 0.04$ ,  $\Delta$ reading time = 96.31 ms, 95% CrI = [0.01, 0.06],  $P(\beta > 0) > .99$  and in

pairwise comparisons with the pronoun-match and pronoun-mismatch conditions individually—pronoun match:  $\hat{\beta} = 0.04$ ,  $\Delta$ reading time = 85.38 ms, 95% CrI = [0.01, 0.06],  $P(\beta > 0) = .99$  pronoun mismatch:  $\hat{\beta} = 0.04$ ,  $\Delta$ reading time = 118.46 ms, 95% CrI = [0.01, 0.08],  $P(\beta > 0) > .99$ . That is, *she* was read slower than *he* even when the reader had already received unambiguous information about gender in the previous sentence. Readers also found it easier to accommodate a switch from first *she* to *he* in the next sentence than the reverse switch. There was no evidence for an interaction of pronoun-match condition and the gender of the second pronoun,  $\hat{\beta} = 0.01$ ,  $\Delta$ reading time = 26.97 ms, 95% CrI = [-0.03, 0.05],  $P(\beta > 0) = .72$  (i.e., their effects were approximately additive).

Although we saw far less change during the preelection period in comprehension than in production, the reading-time advantage for *he* over both *she* and *they* increased significantly in the immediate postelection round relative to the preelection period (Fig. 2c)—increase in *she* disadvantage:  $\hat{\beta} = 0.06$ ,  $\Delta$ reading time = 154.69 ms, 95% CrI = [0.02, 0.11],  $P(\beta > 0) > .99$  increase in *they* disadvantage:  $\hat{\beta} = 0.99$ ,  $\Delta$ reading time = 135.50 ms, 95% CrI = [-0.00, 0.11],  $P(\beta > 0) = .97$ . This result demonstrates that, under some circumstances, changes in world-event expectations can rapidly and dramatically affect comprehension.

To rule out the possibility of intrinsic differences in reading times for *he* and *she*, we conducted an auxiliary experiment ( $N = 1,897$ , described in detail in the Supplemental Material). The auxiliary experiment involved reading of three-sentence texts, similar to those used in Experiment 1, in which a gendered pronoun was encountered where a proper-name antecedent provided near-categorical expectations regarding the referent's gender (e.g., "It was Sunday and the zoo was very crowded. When little [Annabelle/Jimmy] saw the baby elephant, [she/he] shrieked in excitement"), where a gender-balanced (according to a separate norming study) proper name served as antecedent (e.g., "It was Sunday and the zoo was very crowded. When little Morgan saw the baby elephant, [she/he] shrieked in excitement"), and where the pronoun was a cataphor (a pronoun whose referent comes later in the sentence), and no animate referents preceded the pronoun in the context sentence (e.g., "It was Sunday and the zoo was very crowded. When [she/he] saw the baby elephant, little [Annabelle/Jimmy] shrieked in excitement").

In none of these cases, nor in several others we analyzed, did the results provide evidence for a reliable or sizeable intrinsic difference in reading times for *she* versus *he*. Thus the reading-time disadvantage for *she* in our main experiment cannot plausibly be attributed to intrinsically slower reading than for *he* independent of context.

## Experiment 2: The 2017 U.K. General Election

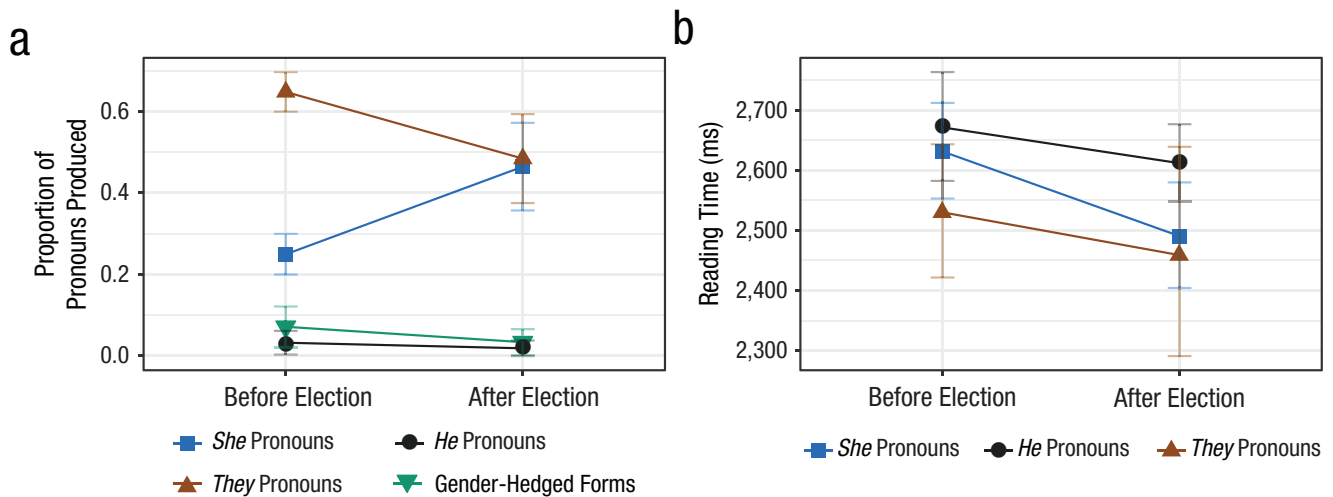
### Method

We conducted an analogous follow-up experiment around the midsummer 2017 U.K. general election, exploiting historical differences between the U.S. and U.K. to investigate the generality of our U.S. results. Whereas the U.S. has always had male heads of government, the U.K. had a female head of government (Margaret Thatcher) from 1979 to 1990; the female sitting prime minister, Theresa May, was strongly expected throughout the campaign to win.

Methods and materials for the text-completion and self-paced reading tasks were the same as in Experiment 1, except where otherwise noted. Regrettably, we did not conduct a belief-estimation task in the preelection round, but the betting odds of May's reelection were well above 80% throughout the campaign (other plausible candidates for prime minister were male), and the general response in the U.K. to May's electoral victory was not surprise that she had won but surprise that the result was as close as it was. Hence, we were confident at the time, and remain confident now, that May was generally considered to have a chance of victory considerably above 50%, which is sufficient for proper interpretation of this experiment. Thus, we chose to assign all our participants to the text-completion and self-paced reading tasks in order to maximize statistical power for those tasks. We did conduct the belief-estimation task in the postelection round, because there appeared to be considerable uncertainty about whether Theresa May would be able to form a government. To account for the chance that a new contender for prime minister would appear on the scene, we allowed participants to enter an arbitrary extra person and to specify that person's chances. Unfortunately, 10% of participants used that slider without specifying a name, so we could not determine the intended person's gender. For this reason, we do not present this data here. However, a detailed analysis is included in the Supplemental Material.

**Participants.** This experiment ( $N = 2,609$ ) consisted of a preelection round (June 5–June 8) and a postelection round (June 17–July 7). We collected data using the U.K. based crowdsourcing platform Prolific (<https://www.prolific.ac/>).

**Apparatus, materials, and procedure.** The same software was used as for Experiment 1. Materials were adapted from Experiment 1 to ensure cultural appropriateness and consistency with the electoral process. The following is an example from the *she* condition:



**Fig. 5.** Results from Experiment 2: (a) rate of pronouns produced and (b) summed reading time before and after the United Kingdom general election. Results are shown separately in (a) for each of the four pronoun types produced and separately in (b) for each of the three pronoun types used in the stimulus sentences. Reading times were summed for the first pronoun and the four subsequent words. Gender-hedged forms include “he or she,” “he/she,” and similar variants. Error bars indicate 95% confidence intervals.

The general election in June 2017 will determine who will serve as prime minister of the UK during the next electoral term. The prime minister will be watched closely when she answers questions during Prime Minister’s Questions. Although the prime minister’s political power will not be absolute, she will wield considerable influence in the government.

After the election, minor changes were introduced into the sentences to ensure consistency with real-world events. See the Supplemental Material for the full list of sentences used in each experimental round. Procedures for the text-completion and self-paced reading tasks were the same as in Experiment 1.

**Data collection and preprocessing.** The first round of data collection took place in the days before and during the election (June 5–June 8, 2017). Data collection stopped when the polling stations closed and first projections were published. The second round of data collection took place after the election when it became increasingly likely that Theresa May would be able to form a government (June 17–July 7, 2017). Data preprocessing was the same as in Experiment 1.

## Results

Overall, the pool of participants was diverse and its composition similar to that tested in Experiment 1. Full participant demographics are reported in Table S5 in the Supplemental Material.

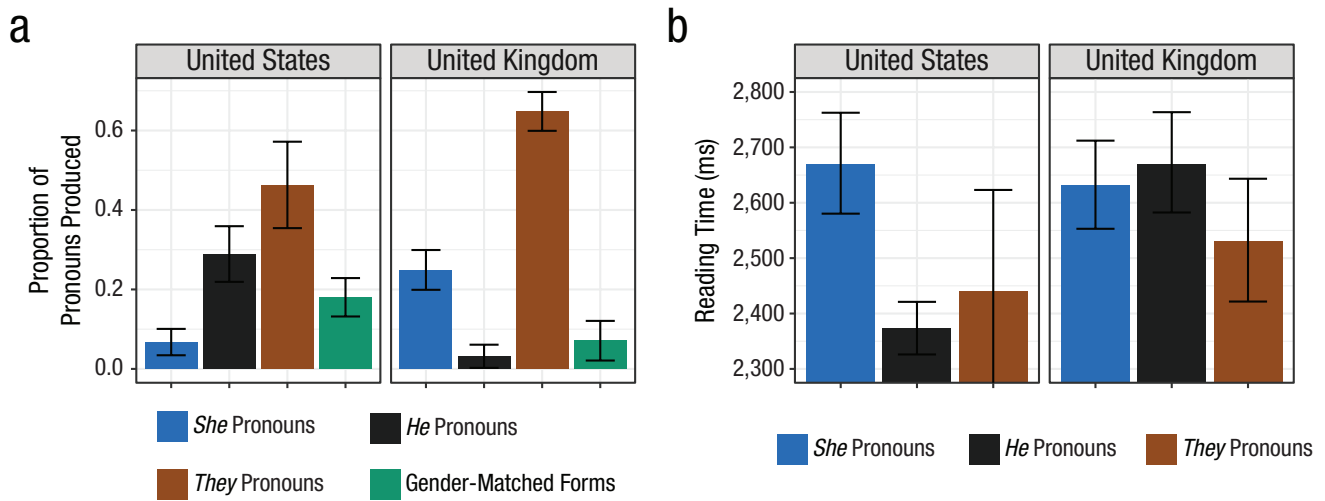
In the text-completion task (Fig. 5a), *they* references were by far the most frequent choice before the

election at 65%, but *they* references dropped to 49% after the election. *She* references jumped from 25% before the election to 47% immediately after the election. *He* pronouns and variants of “he or she” were virtually never used either before or after the election. Compared with the U.S., U.K. usage of *she* and *they* references was considerably higher and *he* and variants of “he or she” were used much less often (Fig. 6a).

In the self-paced reading task (Fig. 5b), initial *she* references were read approximately as slowly as in the U.S. (Fig. 6b). However, *he* elicited similar slowdowns as *she* references,  $\hat{\beta} = -0.02$ ,  $\Delta$ reading time =  $-51.94$  ms, 95% CrI =  $[-0.06, 0.02]$ ,  $P(\beta < 0) = .83$ , in contrast to results in the U.S. Neutral *they* references were read substantially faster than *he* and *she*,  $\hat{\beta} = -0.06$ ,  $\Delta$ reading time =  $-159.18$  ms, 95% CrI =  $[-0.11, -0.01]$ ,  $P(\beta < 0) = .99$ . After the election, both *she*,  $\hat{\beta} = -0.05$ ,  $\Delta$ reading time =  $-132.08$  ms, 95% CrI =  $[-0.10, 0.00]$ ,  $P(\beta < 0) = .97$  and *they*,  $\hat{\beta} = -0.06$ ,  $\Delta$ reading time =  $-167.93$  ms, 95% CrI =  $[-0.14, 0.02]$ ,  $P(\beta < 0) = .95$ , elicited faster reading times than *he* references.

Second pronouns (Fig. 7) that mismatched the gender of the first pronoun elicited substantial slowdowns,  $\hat{\beta} = 0.11$ ,  $\Delta$ reading time =  $294.31$  ms, 95% CrI =  $[0.08, 0.14]$ ,  $P(\beta > 0) > .99$ . There was a small tendency toward a *she* advantage, but this effect was not reliable,  $\hat{\beta} = -0.02$ ,  $\Delta$ reading time =  $-61.80$  ms, 95% CrI =  $[-0.07, 0.02]$ ,  $P(\beta < 0) = .87$ . Again, there was no evidence for an interaction of pronoun gender and pronoun-match condition,  $\hat{\beta} = -0.03$ ,  $\Delta$ reading time =  $-75.06$  ms, 95% CrI =  $[-0.09, 0.04]$ ,  $P(\beta < 0) = .81$ ; their effects were roughly additive.

While these results suggest that biases in the U.K. against *she* were not so strong as to overwhelm effects of event expectations on linguistic preferences, there



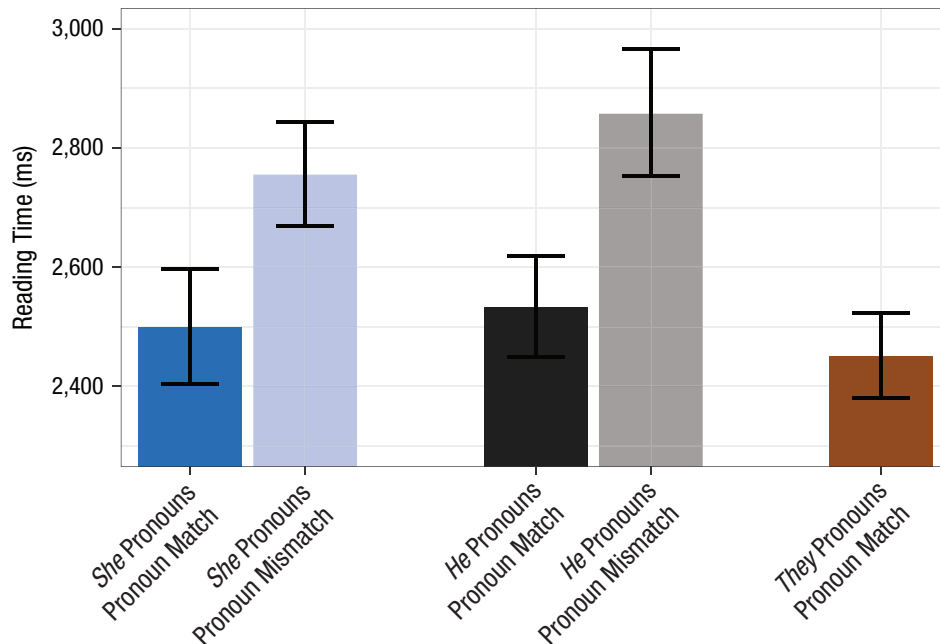
**Fig. 6.** Comparison of pronoun production and comprehension in Experiment 1 (United States election) and Experiment 2 (United Kingdom election): (a) rate of each type of pronoun produced and (b) summed reading time for the first pronoun and the four subsequent words, separately for each type of pronoun read. Gender-hedged forms include “he or she,” “he/she,” and similar variants. Data are from only the preelection periods. Error bars indicate 95% confidence intervals.

are still signs of a *she* pronoun disadvantage in our U.K. comprehension data. In particular, during the preelection round, there was no significant reading-time disadvantage for *he* relative to *she* (Fig. 5b), despite the strong dispreference for *he* in production. A small comprehension disadvantage for *he* emerged in the postelection round (Fig. 5b), but at this point,

the *she*-over-*he* advantage in production was even larger (Fig. 5a).

### Discussion

We found that linguistic preferences can change rapidly as expectations regarding an event change. In both



**Fig. 7.** Results from Experiment 2: summed reading time for the second pronoun and the four subsequent words, separately for each pronoun type and pronoun-match condition. Data are from before and after the election. Means are geometric, and error bars indicate 95% confidence intervals.

countries, the change in event expectation induced by the election itself had a powerful effect on production and comprehension. Our U.S. data, which tracked expectations and linguistic preferences across several months before the election, gave us further insight into differences in how preferences change in production versus comprehension. *He* and *they* production rose and fell in close synchrony with event expectations during the preelection period, whereas comprehension was comparatively unaffected by changes in expectations. We speculate that this greater lability in production might reflect stronger pressure for linguistic preferences to stay up to date with event expectations when the linguistic behavior is easily observable, as in language production, than when the behavior is difficult to observe, as in comprehension.

Largely overwhelming these changes in expectations, however, we found a persistent, severe bias against *she* pronouns relative to expectations about the election outcome. In the case of the U.S. election, this bias took a particularly clear form. In production, expectations that the next president would be male largely manifested as *he* pronoun references, whereas expectations that the next president would be female largely manifested as *they* references, even when the female candidate was expected to win. Likewise, *she* pronouns elicited massive comprehension disruption, even when the female candidate was expected to win. Strikingly, encountering a *she* reference to the next president during reading still did not fully eliminate this bias: When participants encountered a second *she* reference in the next sentence, comprehension was once again disrupted (relative to a second *he* reference in otherwise matched texts). In the U.S., this bias might arise from the president having always been male throughout the country's history. In both countries, this bias might also reflect broader and more general gender biases in society and language use. Our results raise the question of whether this and similar biases may themselves contribute to the spread and persistence of stereotypes in society (Atir & Ferguson, 2018).

Our research goes beyond previous work on gender-stereotype effects in language processing (Banaji & Hardin, 1996; Duffy & Keir, 2004; Foertsch & Gernsbacher, 1997; Kennison & Trofe, 2003; Osterhout et al., 1997; Sturt, 2003) in several respects. First, we are not aware of any previous psycholinguistic study using ongoing world events as the manipulation influencing gender expectations. The success of this natural experiment opens the door to future investigations of how linguistic preferences change in society, analogous to recent studies of other psychological processes, such as social norms and attitudes (Georgeac, Rattan, & Effron, 2018; Tankard & Paluck, 2017).

Second, our observed effect sizes (e.g., a 300-ms reading-time penalty in our U.S. comprehension study) are far larger than those seen in previous research on gender-stereotype effects in language processing (where reading-time penalties typically ranged from 25–100 ms; Duffy & Keir, 2004; Kennison & Trofe, 2003; Sturt, 2003). Our larger effect sizes may reflect the attention commanded by the experimental topic matter for our participants; if this speculation is correct, it would suggest that the effect sizes observed in a wide range of psycholinguistic studies might be substantially lower than the effect sizes that would hold in more ecologically natural contexts.

Third, because our study involved expectations and linguistic preferences regarding specific, high-profile real-world events involving famous people, we were able to measure expectations regarding this event directly, allowing us to test—and ultimately reject—the hypothesis that gender-bias effects on language processing are fully mediated by event expectations. We observed a strong bias against *she* pronouns even among female Democrats during the U.S. electoral campaign (see Fig. S25), a demographic group that expected the female candidate to win throughout our study and that a priori might be least expected to show such a bias. These linguistic biases are all the more striking given the widespread knowledge in the U.S. population that the female candidate, Hillary Clinton, had held some of the most important political posts in the country (e.g., Senator from 2001–2009, Secretary of State from 2009–2013), whereas her general-election opponent, Donald Trump, had never held political office.

It is reasonable to ask whether the disadvantage for *she* pronouns observed in our data might reflect not a biased mapping but rather a default strategy possibly employed by a subset of our experimental participants: When the gender of an event participant is unknown, use *he* to refer to them (generic *he*). It is difficult to ascertain exactly how categorically this possibility can be ruled out; we address the matter at length in the Supplemental Material. But there are two simple reasons why we believe a default *he* strategy is unlikely as a complete explanation of our data. First, in our U.K. experiment, *he* was almost never produced either before or after the election (Fig. 5a). Second, during the U.S. preelectoral period, as expectations strengthened that the female candidate would win and *he* production rates dropped, the production rates that rose were not *she* but rather *they* forms, in all demographic groups, including those in which the female candidate was expected to win throughout our study. That is, the bias we observe in our data is less well characterized as a bias for *he*, as a default strategy would predict, and better characterized as a bias against *she*.

Further clarifying the nature and mechanisms of these implicit linguistic biases is a matter for future research. One possible mechanism could be that phrases such as “U.S. president” prime male-oriented language more than female-oriented language because of the co-occurrence of male-oriented language and language related to U.S. presidents in past experience. If true, these biases would follow the template definition of implicit biases, as proposed in implicit-cognition theory (Greenwald & Banaji, 1995). A consequence would be that even when our beliefs about the world are updated quickly, our linguistic preferences may be guided by past experience and thus take much longer to adapt.

Pronoun preferences in English are a topic of perennial interest, ranging from controversy over singular *they* (Baron, 1981; Foertsch & Gernsbacher, 1997; Freeman, 2017) to the more recent rising awareness of nonbinary gender identity and individuals’ preferred pronouns (Chak, 2015). Our data suggest that singular *they* is unremarkable and even preferred by many of our participants, especially younger ones (see Fig. S28). But this article reports a previously undocumented powerful bias affecting linguistic behavior: When people refer to the next head of government, even when the set of candidates has been narrowed down to two, they disprefer *she* references relative to expectations regarding the referent’s gender. More broadly, our results demonstrate the potential of psycholinguistic methods for investigations of how linguistic behavior evolves over time and for revealing and deepening our understanding of implicit cognition.

### Action Editor

Rebecca Treiman served as action editor for this article.

### Author Contributions

T. Poppels and T. von der Malsburg developed the initial idea for this study and ran the two pilot experiments. T. von der Malsburg, T. Poppels, and R. P. Levy jointly designed the study. Data were collected by T. von der Malsburg and T. Poppels. T. von der Malsburg analyzed the data. T. von der Malsburg and R. P. Levy wrote the article and Supplemental Material with help from T. Poppels.

### ORCID iD

Titus von der Malsburg  <https://orcid.org/0000-0001-5925-5145>

### Acknowledgments

We are grateful to Veronica Boyce, Evelina Fedorenko, Edward Gibson, Nancy Kanwisher, Josh McDermott, Laura Niemi, Rebecca Saxe, Laura Schulz, and Stephanie Smith for comments on previous versions of the manuscript. We are

also grateful to Craig Thorburn, Emily Troscianko, Susan Blackmore, and Nicholas Cooper for their advice in adapting our materials for the United Kingdom election.

### Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

### Funding

T. von der Malsburg was funded through a Feodor Lynen Research Fellowship by the Alexander von Humboldt Foundation and through National Institutes of Health Grant No. HD065829, awarded to R. P. Levy and K. Rayner. R. P. Levy gratefully acknowledges support from an Alfred P. Sloan Fellowship.

### Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797619890619>

### Open Practices



All data, analysis scripts, and study materials have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/gx5tr/>. The design and analysis plans for the experiments were not preregistered. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797619890619>. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.

### Note

1. Further, in our U.K. general election study (Experiment 2), where a female candidate was expected to win, *she* was used far more often than *he*, which is inconsistent with the idea that *he* is a generic default in English.

### References

- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.
- Atir, S., & Ferguson, M. J. (2018). How gender determines the way we speak about professionals. *Proceedings of the National Academy of Sciences, USA*, *115*, 7278–7283.
- Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science*, *7*, 136–141.
- Baron, D. E. (1981). The epicene pronoun: The word that failed. *American Speech*, *56*, 83–97.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep

- it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63, 489–505.
- Bureau of Labor Statistics. (2018). *Labor force statistics from the current population survey*. Washington, DC: U.S. Department of Labor.
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356, 183–186.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1). doi:10.18637/jss.v076.i01
- Casey, L. S., Chandler, J., Levine, A. S., Proctor, A., & Strolovitch, D. Z. (2017). Intertemporal differences among MTurk workers: Time-based sample variations and implications for online data collection. *SAGE Open*, 7(2). doi:10.1177/2158244017712774
- Chak, A. (2015). Beyond 'he' and 'she': The rise of non-binary pronouns. *BBC News*. Retrieved from <http://www.bbc.com/news/magazine-34901704>
- Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, 32, 551–559.
- Foertsch, J., & Gernsbacher, M. A. (1997). In search of gender neutrality: Is singular *they* a cognitively efficient substitute for generic *he*? *Psychological Science*, 8, 106–111.
- Franke, M., Dablander, F., Schoeller, A., Bennett, E., Degen, J., Tessler, M. H., . . . Goodman, N. D. (2016). What does the crowd believe? A hierarchical approach to estimating subjective beliefs from empirical data. In A. Papafragou, D. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 2669–2674). Austin, TX: Cognitive Science Society.
- Freeman, J. (2017). A letter from an Ivy League admissions dean. *The Wall Street Journal*. Retrieved from <https://www.wsj.com/articles/a-letter-from-an-ivy-league-admissions-dean-1492107041?mg=prod/accounts-wsj>
- Garnham, A., Doehren, S., & Gyax, P. (2015). True gender ratios and stereotype rating norms. *Frontiers in Psychology*, 6, Article 1023. doi:10.3389/fpsyg.2015.01023
- Georgeac, O. A. M., Rattan, A., & Effron, D. A. (2018). An exploratory investigation of Americans' expression of gender bias before and after the 2016 presidential election. *Social Psychological and Personality Science*, 10, 632–642.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27.
- Greenwald, A. G., & Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California Law Review*, 94, 945–967.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- He, W., Goodkind, D., & Kowal, P. (2016). *An aging world: 2015*. Washington, DC: U.S. Government Publishing Office.
- Howden, L. M., & Meyer, J. A. (2011). *Age and sex composition: 2010*. Washington, DC: U.S. Government Publishing Office.
- Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review*, 100, 109–128.
- Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, 32, 355–378.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31, 32–59.
- Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *SAGE Open*, 6(1). doi:10.1177/2158244016636433
- Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189, 226–228.
- Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 69–89). Hillsdale, NJ: Erlbaum.
- Mitchell, D. C. (2004). On-line methods in language processing: Introduction and historical review. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension* (1st ed., pp. 15–32). New York, NY: Psychology Press.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18, 1098–1111.
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25, 273–285.
- Real Clear Politics. (2016). *U.S. presidential election - betting odds*. Retrieved from [https://www.realclearpolitics.com/epolls/writeup/us\\_presidential\\_election\\_live\\_betting\\_odds.html](https://www.realclearpolitics.com/epolls/writeup/us_presidential_election_live_betting_odds.html)
- Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128, 302–319.
- Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing samples in cognitive science. *Trends in Cognitive Sciences*, 21, 736–748.
- Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Tankard, M. E., & Paluck, E. L. (2017). The effect of a Supreme Court decision regarding gay marriage on social norms and personal attitudes. *Psychological Science*, 28, 1334–1344.
- Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Topics in Cognitive Science*, 18, 605–611.
- Wang, S. (2016). *Final projections 2016*. Retrieved from <http://election.princeton.edu/2016/11/08/final-mode-projections-clinton-323-ev-51-di-senate-seats-gop-house/>