





















REPLICATION STUDY 

The effects of exposure and explicit stereotypes on veracity judgments of Polish-accented English speech: A preregistered close replication and extension of Boduch-Grabka & Lev-Ari (2021)

Samantha Barlow , Greg Beardsley , Zéta Bsharah, Robin Crofts , Carlos De La Rosa , Andrea Gutierrez , Carlie Highfill , Amy Gail Wade Johnson, Caroline Johnson , Jacob Johnson , Isaac Leyva Cardenas , Jordan Taylor Martinez , Nathaniel Todd Miller , Riley Monroe Murray , Sylvia Page , Taylor Petersen , Irina Ramos , Rayvin Rhodes , Phoebe Vainuku , Brenan M. Wednesday , Emma Corrine Farnsworth , Seung Kyung Kim and Rachel Hayes-Harb 

University of Utah

Corresponding author: Rachel Hayes-Harb; Email: r.hayes-harb@utah.edu

(Received 29 June 2023; Revised 05 December 2023; Accepted 11 January 2024)

Abstract

Boduch-Grabka and Lev-Ari (2021) showed that so-called “native” British-English speakers judged statements produced by Polish-accented English speakers as less likely to be true than statements produced by “native” speakers and that prior exposure to Polish-accented English speech modulates this effect. Given the real-world consequences of this study, as well as our commitment to assessing and mitigating linguistic biases, we conducted a close replication, extending the work by collecting additional information about participants’ explicit biases towards Polish migrants in the UK. We did not reproduce the original pattern of results, observing no effect of speaker accent or exposure on comprehension or veracity. In addition, the measure of explicit bias did not predict differential veracity ratings for Polish- and British-accented speech. Although the current pattern of results differs from that of the original study, our finding that neither comprehension nor veracity were impacted by accent or exposure condition is not inconsistent with the Boduch-Grabka and Lev-Ari (2021) processing difficulty account of the accent-based veracity judgment effect. We explore possible explanations for the lack of replication and future directions for this work.

Please see the authors’ positionality and contribution statements following the recommendations of Allen et al. (2019) at <https://osf.io/etgc4>.

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Introduction

A number of studies have demonstrated that so-called “native” and “non-native” speakers judge statements produced by “foreign-accented” speakers as less likely to be true than those produced by “native speakers” (e.g., Boduch-Grabka & Lev-Ari, 2021; Hanzlíková & Skarnitzl, 2017; Lev-Ari & Keysar, 2010). Lev-Ari and Keysar (2010) further found that raising listeners’ awareness to the (presumed) source of difficulty by asking them to rate the speech for understandability partially modulated the negative effect, leading the authors to conclude that these judgments of reduced veracity of “non-native” speech can be attributed to processing difficulty. Other studies have similarly investigated listeners’ judgments of the veracity of statements produced by “native” and “non-native” speakers, but with mixed results. Some have observed no reduction in veracity judgments for “non-native” speech (e.g., Souza & Markman, 2013; Wetzel & Gygas, 2021), even when their listeners show evidence of holding stereotypes about the credibility of “non-native” speakers in a separate task (e.g., Stocker, 2017). Importantly, there is debate in this literature concerning whether observed reductions in veracity judgments for “non-native” speech are attributable to *processing difficulty*, *accent-based prejudice*, or *both* (e.g., Castillo *et al.*, 2014). These studies have varied along many dimensions, including geographic and national context, languages, speech content, and the type of rating task, leading to crucial questions about the “nature, repeatability, and generalizability” of the effect (Liu *et al.*, 2023).

Study to be replicated

Building on this growing line of research, Boduch-Grabka and Lev-Ari (2021) sought to further isolate the impact of processing difficulty on veracity judgments by easing this difficulty through exposure to “native” or “non-native” accents prior to the veracity judgment task. Studies have shown that comprehension of unfamiliarly accented speech improves following even a few seconds of exposure (e.g., Clarke & Garrett, 2004), and Boduch-Grabka and Lev-Ari (2021) hypothesized that if the “[veracity judgment] effect is at least partly due to difficulty of processing the speech, then improving listeners’ ability to understand the speech should reduce their tendency to find the speech less credible” (p. 5). In their study, 220 “native speakers” of English were randomly assigned to “British” (i.e., “native” British English speech) or “Polish” (i.e., “Polish-accented” English speech) exposure conditions. During the exposure phase, they listened to eight brief stories in English ($M = 179$ words), each read aloud by a different “British” or “Polish” speaker, according to exposure condition. Next, participants listened to one of two counterbalanced lists of 50 trivia statements, half produced by eight “Polish” speakers and the other half by six “British” speakers; half of the statements were true, and half were false. Participants judged each statement on a 100-point continuous false–true scale. Finally, all participants heard eight sentences extracted from the exposure phase stories produced by the “Polish” speakers and transcribed them. They found that (1) participants in the “British” exposure condition judged “British-accented” trivia statements as more true than those produced by “Polish” speakers; (2) there was an interaction between exposure and trivia statement accent such that Polish-accented exposure reduced the difference in veracity judgments of British-accent versus Polish-accented speech; (3) “Polish” exposure participants judged Polish-accented trivia statements as more true than did “British” exposure participants; (4) “Polish” exposure participants were more accurate than “British” exposure participants at transcribing the Polish-accented sentences in the

comprehension task; and (5) a mediation test demonstrated that once comprehension score was taken into account, the effect of exposure condition was no longer significant, a finding the authors interpret as indicating that “the reason that the exposure to Polish accent increased belief in statements delivered in Polish-accented speech is because it improved participants’ comprehension of the accent” (p. 9). Boduch-Grabka and Lev-Ari (2021) concluded that the prior exposure to “Polish-accented” speech led to more accurate comprehension for participants in the “Polish” exposure condition, which in turn moderated the negative effect of “non-native” speech on veracity judgments.

The current study

The current study advances the authors’ commitment to interrogating and moderating the socio-cognitive biases that underpin responses to language and language users (please see the author positionality statement at <https://osf.io/etgc4>). Research on the perceived veracity of speech produced by speakers from various backgrounds has important implications for the fields of applied linguistics, criminal justice, psychology, and many others. If, as demonstrated by Boduch-Grabka and Lev-Ari’s (2021) study, accent-based veracity effects can be mitigated via minimal interventions (in this case, by brief prior experience with the “accented” speech), replication studies should be conducted to confirm and expand upon the findings. The Boduch-Grabka and Lev-Ari (2021) study is thus an excellent candidate for independent replication, and the current study allows us to determine the replicability of the Boduch-Grabka and Lev-Ari (2021) findings.

The preregistration for this replication study can be found at: <https://osf.io/ry8hm>, and study materials, data, and analysis code are available at: <https://osf.io/etgc4>. This is a “close replication” of Boduch-Grabka and Lev-Ari (2021) following the characterization provided by Porte and McManus (2019, starting p. 72). A significant modification to Boduch-Grabka and Lev-Ari’s (2021) study procedures is the introduction of a new continuous participant background variable: an index of explicit bias towards Polish migrants in the UK (adapting the explicit bias task of Babel and Russell, 2015). Given the mixed results and relatively small effect sizes associated with earlier studies—along with the debate over the source of the veracity effects that have been found—we hypothesized that this variable may help to clarify the factors that control the effect size, source, and generalizability of the findings. The addition of the explicit bias variable allows us to examine the independent contributions of *processing difficulty* and *accent-based prejudice* on veracity judgments. As suggested by Boduch-Grabka and Lev-Ari (2021), “[i]t is quite possible that participants’ lower belief in the Polish-accented statements was due to both prejudice and processing difficulty” (p. 11); here we ask whether we can account for additional variance in the veracity judgment data by adding a measure of accent-based prejudice.

Terminology

We have placed quotation marks around the terms “native” and “non-native” when referencing their use in previous studies to draw attention to the problematic nature of these terms. Cheng et al. (2021) note that “simply reporting that ‘native speakers’ participated or recorded stimuli clearly does not provide information adequate for replication” (p. 7) and that unacknowledged differences in the definitions of the term “native” and “non-native” among researchers may even be partially responsible for low

rates of replication in psychology and related fields. We affirm the problematic nature of such essentializing terms, which lead to vagueness and potential harm in psycholinguistics research. Thus, an additional change from the original study is that we collected and reported richly detailed information from our study participants about their language backgrounds using instruments recommended by Cheng *et al.* (2021).

We additionally acknowledge that the characterization of study materials as representing “British accent” and “Polish accent” is similarly problematic: Is the English produced by Polish migrants in Britain not also a variety of British English? How were the individuals who represented these accents selected for the original study’s materials? Because we employ the very same speech materials as the original study and do not know the precise criteria that were used to select the speakers, we reproduce to some extent the vagueness and potential harm associated with these labels. However, to be more precise about what we assume to be the relevant properties of the speech samples and to acknowledge that all materials are produced by speakers of English, we use the terms “Polish-accented English” and “British-accented English” when referring to speech materials, and “Polish-accented” and “British-accented” to refer to the speakers who produced the materials.

Method

Participants

Participants ($N = 222$, following Boduch-Grabka and Lev-Ari’s (2021) sample size of 220) were recruited via [Prolific.com](https://prolific.com) and were paid the equivalent of 20 USD/hour for completing the ~28-minute study. In the original study, “[p]articipants were first screened for native language and having no Polish friends or family members” (p. 6). On the assumption that “native language” refers to British English, we used the following Prolific screening criteria to recruit the sample of 222: country of birth, nationality, and current location were identified as the United Kingdom; current UK area of residence was England; and first language, earliest language in life, and primary language were English (these are the demographic category options provided by Prolific that were best-suited to the goal of recruiting “native speakers” of British English). We used a post-experiment questionnaire administered via [Qualtrics.com](https://qualtrics.com) to collect additional information about these participants and confirmed that none identified “Polish people” as their predominant social group. Two participants reported problems when registering their responses (one was unable to type responses during the comprehension task, and the other encountered an error while completing the participant questionnaire), and four participants did not meet the original study’s criterion of achieving 6/8 accuracy on the attention-check questions (“[t]o be included, participants had to respond correctly to at least six of the eight [attention check] questions” (p. 7); see below for attention-check task details). All 216 remaining participants reported that they consider English to be (one of) their native language(s). Many reported having studied and/or being familiar with one or more additional languages: Ancient Greek ($n = 1$), Arabic ($n = 6$), British Sign Language ($n = 5$), Cantonese ($n = 2$), Danish ($n = 1$), Dutch ($n = 7$), Egyptian ($n = 1$), French ($n = 93$), German ($n = 58$), Greek ($n = 3$), Gujarati ($n = 1$), Hindi ($n = 2$), Icelandic ($n = 1$), Igbo ($n = 1$), Irish ($n = 1$), Italian ($n = 17$), Japanese ($n = 11$), Korean ($n = 5$), Ladino ($n = 1$), Latin ($n = 13$), Lingala ($n = 1$), Mandarin ($n = 1$), Polish ($n = 4$), Portuguese ($n = 3$), Russian ($n = 5$), Spanish ($n = 57$), Swedish ($n = 2$), Tagalog ($n = 1$), Turkish ($n = 1$), Twi ($n = 1$), Urdu ($n = 3$),

Table 1. Summary of notable participant comparisons between Boduch-Grabka and Lev-Ari (2021) and the current study

Boduch-Grabka and Lev-Ari (2021)	Current study
“[R]ecruited via contacts, mailing lists, and social media” (p. 6).	Recruited via Prolific.com
“They were often holders of public sector roles (e.g., police personnel)” (p. 6); however, proportion of participants who were public sector role holders is unspecified in the article	Public sector employees not specifically targeted
Compensation unspecified	Compensated at equivalent of US\$20/hr
220 participants; all met criterion for attention check items	222 participants; 4 of these excluded for attention check failure and 2 excluded for technical problems; see below for analyses involving additional exclusionary criteria.
“[S]creened for native language” (p. 6); presumably they were asked if they were “native speakers” of British English	First language, earliest language in life, and primary language were English
$F = 124$; Age range: 18–74, $M = 38.15$, $SD = 14.36$ (p. 6)	121 participants identified as female, 92 as male, 1 as “neither of these”, and 2 selected “I prefer not to say”; Age range: 19–87, $M = 40.0$, $SD = 13.02$
Country of birth, nationality, current location, and current residence not specified.	Country of birth, nationality, and current location was the UK, and current residence was England.
“[N]o Polish friends or family members” (p. 6).	Identified “British people” or “Other” (not “Polish people”) as the predominant ethnic composition of their social group, including friends, family, and co-workers.

Welsh ($n = 6$), and Yoruba ($n = 2$). Their Prolific profiles indicated that they came from the following regions within the UK: East Midlands ($n = 20$), East of England ($n = 28$), London ($n = 29$), North East ($n = 12$), North West ($n = 29$), South East ($n = 27$), South West ($n = 22$), West Midlands ($n = 26$), and Yorkshire and the Humber ($n = 22$). One participant met the location criterion but did not give consent to report the specific region. A fuller report of participant characteristics can be found in the OSF repository. [Table 1](#) summarizes notable recruitment and participant differences between the original and current studies.

Materials

All audio materials used in the present study were those used by Boduch-Grabka and Lev-Ari (2021) and were retrieved from their Open Science Framework repository (<https://osf.io/a2jcw/>). Additional materials, including the attention-check task questions, some experiment instructions, and information about which sentences to extract from the exposure audio files for use in the comprehension task, were secured via email communication with original study author Lev-Ari. All other materials were inferred from the Boduch-Grabka and Lev-Ari (2021) article (e.g., some experiment

instructions, participant background questions, and experiment presentation code) or created by the present authors (e.g., the explicit bias task statements), and can be retrieved at <https://osf.io/etgc4/>.

Procedure

All data reported here were collected online during November 2022. Participants were required to use a laptop or desktop computer with a keyboard and headphones to participate in the online experiment (developed using Psychopy (Peirce *et al.*, 2019), and hosted online via [Pavlovia.org](https://pavlovia.org)) and the accompanying post-experiment questionnaire and explicit bias task (hosted online via [Qualtrics.com](https://qualtrics.com)). Upon consenting to participate and completing a four-question sound check task (multiple-choice auditory word identification administered via [Qualtrics.com](https://qualtrics.com)) with 100% accuracy, participants were randomly assigned to one of two exposure conditions (British-accented exposure or Polish-accented exposure) and to one of two counterbalancing list conditions (List 1 or List 2) and were automatically redirected to the experiment's tasks.

Exposure phase

Participants listened to eight randomly ordered audios of paragraph-length statements, each produced by one of eight Polish-accented English speakers or one of eight British-accented English speakers, according to the participant's exposure condition. Prior to listening to the statements, participants read the following paragraph (adapted from materials received from author Lev-Ari via email correspondence):

Police personnel are often trained in how to better understand and evaluate statements made by victims and witnesses. We are studying how members of the general public understand and evaluate such statements. We would therefore ask you to listen to both police-related and neutral statements and ask you questions about them. According to recent statistics, Poles are the biggest non-UK-born population in the UK with around 853,000 Poles residing in the UK in 2016. For this reason, the recordings you will listen to might include several Polish speakers.

After hearing each passage, participants responded to a multiple-choice listening comprehension test item to confirm that they were paying attention, for a total of eight attention-check items (also received from author Lev-Ari via email correspondence).

Veracity judgment task

Next, participants judged the veracity of 50 randomly ordered trivia statements (half of which were true and half of which were false), each of which was produced by one of eight Polish-accented English speakers or one of six British-accented English speakers. Half of the statements were Polish-accented and the other half British-accented, with the language background of the speakers counterbalanced across List 1 and List 2. Following each statement, participants were asked to use a FALSE–TRUE slider to indicate how likely they thought the statement to be true. There was no time limit, and participants advanced to the next item immediately upon registering a response.

Comprehension task

One sentence from each of the eight Polish-accented witness statements was extracted for presentation in the comprehension task (the identity of these sentences was provided by author Lev-Ari via email correspondence). Sentences were presented in random order, and following the presentation of each sentence, participants were asked to transcribe the sentence by typing in a text box. There was no time limit, and participants advanced to the next item immediately upon pressing the enter/return key.

Participant background questionnaire

Upon completing the comprehension task, participants were automatically redirected to a questionnaire to complete the remaining tasks. Participants were prompted to indicate their age and gender, identify the predominant ethnic composition of their social group (“Polish people”, “British people”, or “Other”), and list and describe their relationship to all of the languages they use, following recommendations for robust language background descriptions provided by Cheng et al. (2021).

Explicit bias task

The explicit bias task occurred at the very end of the study session so as not to interfere with the close replication of the original study. The task was adapted from one described by Babel and Russell (2015), which assessed the degree of participants’ explicit stereotyped views of “Asian Canadians” and “White Canadians” in Canada. Their task involved ten statements, half of which would elicit a *Strongly Disagree* (1) response and half of which would elicit a *Strongly Agree* (7) response if participants held stereotyped views about “Asian Canadians”. To create our analogous six-item explicit bias task concerning Polish migrants in the UK, we relied on scholarly literature and media reports of stereotypes about Polish migrants in the UK. Stereotypes included having a strong work ethic (Dunin-Wasowicz, n.d.), taking jobs from British workers (Rzepnikowska, 2019), being “benefits spongers” (Dunin-Wasowicz, n.d.; Portas, 2018), achieving low education levels (Portas, 2018), and having poor English-language skills and unsophisticated manners (Portas, 2018). Table 2 presents the six statements used in the explicit bias task.

On completion of the explicit bias task, participants were automatically redirected to the Prolific site where they were notified that they had received compensation for their

Table 2. Explicit bias task statements

<p>Strongly Disagree (1)</p> <ul style="list-style-type: none"> • Polish migrants in England speak English as well as English people do. • English people are more likely to take advantage of benefits and welfare services than Polish migrants. • Polish migrants typically have more sophisticated manners than do English people. <p>Strongly Agree (7)</p> <ul style="list-style-type: none"> • Polish migrants take jobs away from English people. • English people are more educated than Polish migrants. • Polish migrants have a stronger work ethic than English people.
--

Note: This table presents the six explicit bias task statements, organized by the expected responses if one holds stereotyped views about Polish migrants in England.

Table 3. Summary of notable methodological comparisons between Boduch-Grabka and Lev-Ari (2021) and the current study

Boduch-Grabka & Lev-Ari (2021)	Current study
Data collection date unspecified; prior to February 2021 article submission.	Data collected in November 2022
“The study was conducted online” (p. 6).	The study was conducted online via Qualtrics and PsychoPy/Pavlovia.
Instructions provided to participants generally unspecified.	Task instructions inferred from description of the tasks in the original article and adapted from those provided by the authors via personal correspondence.
Duration of the study not reported	The study took approximately 28 minutes to complete.
No explicit bias task	Explicit bias task

participation. Table 3 summarizes notable material and procedure comparisons between the original and current studies.

Hypotheses

The following hypotheses summarize the findings of Boduch-Grabka and Lev-Ari (2021):

- **Hypothesis 1:** Veracity ratings would be lower overall for Polish-accented statements than for British-accented statements.
- **Hypothesis 2:** We would observe an interaction of Exposure condition and Trivia Speaker condition such that the effect of Trivia Speaker would be smaller for participants in the Polish-accented Exposure condition.
- **Hypothesis 3:** Participants in the Polish-accented Exposure condition would assign higher veracity judgments to Polish-accented statements than would participants in the British-accented Exposure condition.
- **Hypothesis 4:** Participants in the Polish-accented Exposure condition would show more accurate comprehension of Polish-accented English speech samples than would those exposed to British-accented speech.
- **Hypothesis 5:** Comprehension accuracy would positively predict veracity judgments of Polish-accented statements (this is a precondition for the mediation analysis associated with Hypothesis 6).
- **Hypothesis 6:** A mediation analysis would reveal that a significant proportion of the effect of Exposure on the veracity judgments of Polish-accented statements would be due to Comprehension, and that once the effect of Comprehension was taken into account, the effect of Exposure on veracity judgments would no longer be significant.

Hypothesis 1 relates to the finding reported in Boduch-Grabka and Lev-Ari (2021) and elsewhere that veracity ratings are overall lower for “non-native”-accented statements than for “native”-accented statements. Hypotheses 2, 3, and 4 relate to the innovation in the Boduch-Grabka and Lev-Ari (2021) study investigating whether prior exposure moderates an accent-based veracity effect. Hypotheses 5 and 6 concern the mechanism behind any moderating effect of exposure. We further hypothesized that additional

variance in veracity judgments would be accounted for by the inclusion of an index of Explicit Bias such that participants reporting greater belief in stereotypes about Polish migrants in the UK would exhibit lower judgments of truthfulness of Polish-accented statements.

Data coding and analysis procedures

Each of the eight attention check questions was scored for accuracy (1 = correct; 0 = incorrect), for a maximum possible score of 8. As indicated above, participants who did not achieve a score of at least 6 (out of 8) on this task were excluded from the analysis. As in the original study, responses to the veracity judgment items were converted to 0–100 (false–true) scores. The eight comprehension task sentences contained 60 content words. As in the original study, we counted the total number of these words that were correctly transcribed, for a maximum comprehension score of 60. We created a Python script to do this computation and counted only exact matches as correct (the original study does not specify the criterion for matches). The explicit bias scores were coded such that larger values on the 1–7 scale were associated with higher degrees of explicit bias.

As in the original paper, we used mixed-effects models for statistical analyses. All analyses were carried out using R's lme4 and lmerTest packages. Participants and items were specified as random factors. We report the results from models with the maximal random effects structure justified by the data, consistent with the original paper's approach. See below for more detailed information about model specifications.

Results

To test the effects of exposure and speaker accent on veracity judgements, we created a mixed-effects model with veracity ratings (0–100 scale) as the dependent variable; Exposure condition (British-accented exposure, Polish-accented exposure), Trivia Speaker condition (British-accented statement, Polish-accented statement), and the interaction of the two as fixed effects. The truth value of each trivia statement, trial number, and list number were included as control factors. The model reported below includes random intercepts for participants and items, and by-participant random slopes for trivia speakers. Categorical variables were sum-coded.¹

The model showed that there was no main effect of Trivia Speaker ($\beta = 0.41$, $SE = 1.279$, $t = 0.32$, $p = 0.75$), no main effect of Exposure ($\beta = 0.37$, $SE = 0.45$, $t = 0.83$, $p = 0.41$), and no interaction between Exposure and Trivia Speaker ($\beta = 0.21$, $SE = 0.29$, $t = 0.74$, $p = 0.46$). Thus, we did not find the effects of Trivia Speaker, Exposure, or the interaction of the two reported by Boduch-Grabka and Lev-Ari (2021). Our own descriptive analyses of the original study's data produced the following means for veracity judgments: in the British-accented Exposure condition, Polish-accented Trivia Speakers = 38.23 and British-accented Trivia Speakers = 60.48;

¹The mixed-effect model specification of Boduch-Grabka and Lev-Ari (2021) includes the same independent variable, the same fixed effect factors, and the same random effect structure. Our specification is slightly different from theirs as we included control factors and employed sum-coded variables instead of treatment-coded variables. These modifications were to improve general model fit and interpretability, and they did not change the structure of the variables of interest. Running the analysis with the exact model specification of Boduch-Grabka and Lev-Ari (2021) returned the same results.

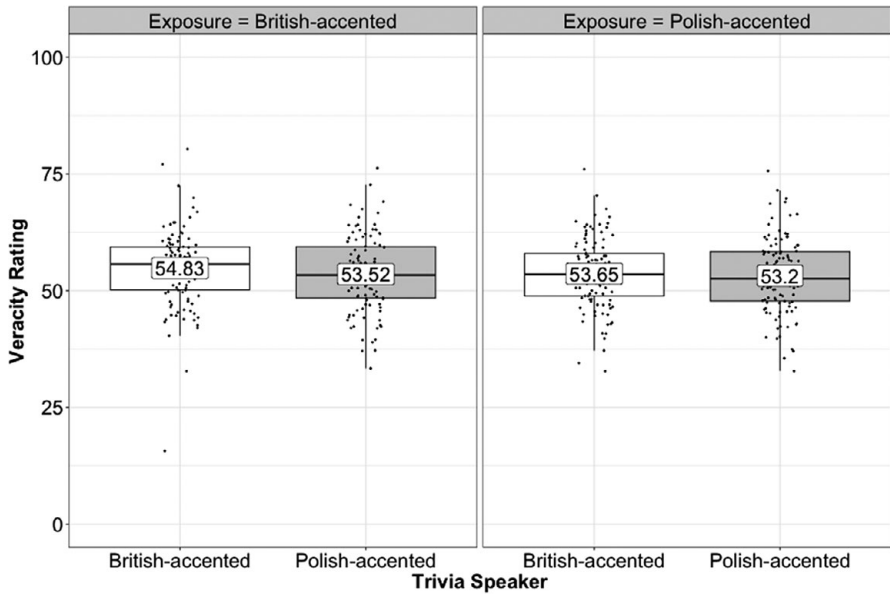


Figure 1. Box plot of veracity ratings by Trivia Speaker condition.

Note: Data from participants in the British-accented exposure condition are presented in the left panel; Polish-accented exposure condition in the right panel. Dots indicate individual participant means, and medians are presented as horizontal lines. Superimposed values show the group mean for the corresponding condition.

in the Polish-accented Exposure condition, Polish-accented Trivia Speakers = 43.77 and British-accented Speakers = 57.96. As can be seen in [Figure 1](#), the means in all four conditions in the current study ranged from 53.20 to 54.83, showing no evidence of the Trivia Speaker and Exposure effects observed in the original study, and thus providing no support for Hypotheses 1, 2, or 3.

In a departure from the analyses conducted by the original authors, to accommodate the possibility that participants did not use the full range of veracity ratings, which could mask the effects of Trivia Speaker and Exposure, we z-scored the veracity ratings of each participant and created a mixed-effects model with the same specifications presented above. The random effects structure for this model included random intercepts for items. The results of modeling this z-scored data were the same. There was no main effect of Trivia Speaker ($\beta = 0.01$, $SE = 0.05$, $t = 0.2$, $p = 0.82$), no main effect of Exposure ($\beta = -1.13e-17$, $SE = 0.008$, $t = 0$, $p = 1$), and no interaction between Exposure and Trivia Speaker ($\beta = -0.005$, $SE = 0.008$, $t = -0.3$, $p = 0.57$).

As in the original study, we performed a linear regression with Comprehension Score as the dependent variable and Exposure condition (British-accented exposure, Polish-accented exposure) as a predictor to test if exposure to Polish-accented English speech was associated with higher comprehension scores of Polish-accented English sentences. Our own analyses of the original study's data revealed that the British- and Polish-accented exposure conditions had mean comprehension scores of 35.99 and 51.71, respectively. In contrast, with means of 50.47 (British-accented exposure) and 50.10 (Polish-accented exposure; see [Figure 2](#)), the current study showed no difference in transcription accuracy between the two Exposure conditions ($\beta = -0.37$,

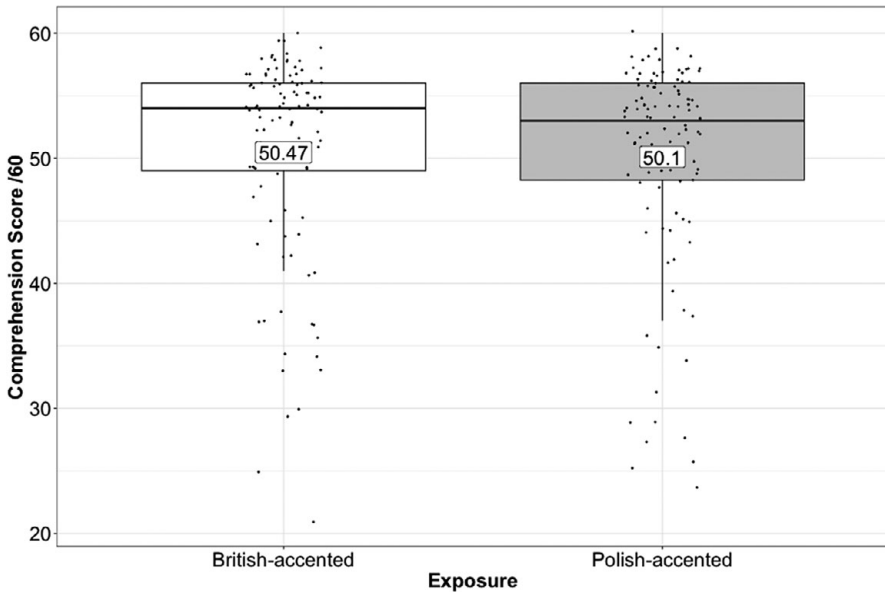


Figure 2. Box plot of comprehension task scores by Exposure condition.

Note: Data from participants in the British-accented Exposure condition are on the left; Polish-accented Exposure condition on the right. Dots indicate individual participant means, and medians are presented as horizontal lines. Superimposed values show the group mean for the corresponding condition.

$SE = 1.13$, $t = -0.33$, $p = 0.74$), and thus no support for Hypotheses 4 or 5. Because this pattern of results did not meet the criteria for a mediation analysis, we did not conduct the subsequent mediation analyses reported in the original study and associated with Hypothesis 6.

The above analyses are based on data from 216 participants, with exclusions based only on criteria explicitly mentioned in the original article. There are, however, additional exclusionary criteria employed by researchers to ensure data quality. For an experiment conducted remotely, it is common to exclude participants who took too long to complete the task, as it suggests participants might have been distracted or experienced disruptions. In the following analyses, we excluded an additional 15 participants who took longer than 40 minutes to complete the task. In addition, because we collected rich language histories of our participants, we were able to identify and exclude data from four participants who reported prior study or other experience with the Polish language. Finally, we excluded data from nine participants who did not identify “British people” as their predominant social groups. Twenty-seven participants met one or more of these additional exclusionary criteria, resulting in a data set of 189 participants. When we repeated the analyses described on this smaller yet cleaner dataset, the results with respect to the hypotheses were the same. There were no effects of Exposure or Trivia Speaker on veracity ratings. There was no effect of Exposure on comprehension scores. Given the lack of differences in the findings based on the two data sets, we conducted the remaining analyses with this smaller and cleaner data set.

Next, we conducted (pre-registered) exploratory analyses to determine whether variance in the veracity judgment data was accounted for by the continuous index of explicit bias that was collected via the explicit bias task. To this end, we created a mixed

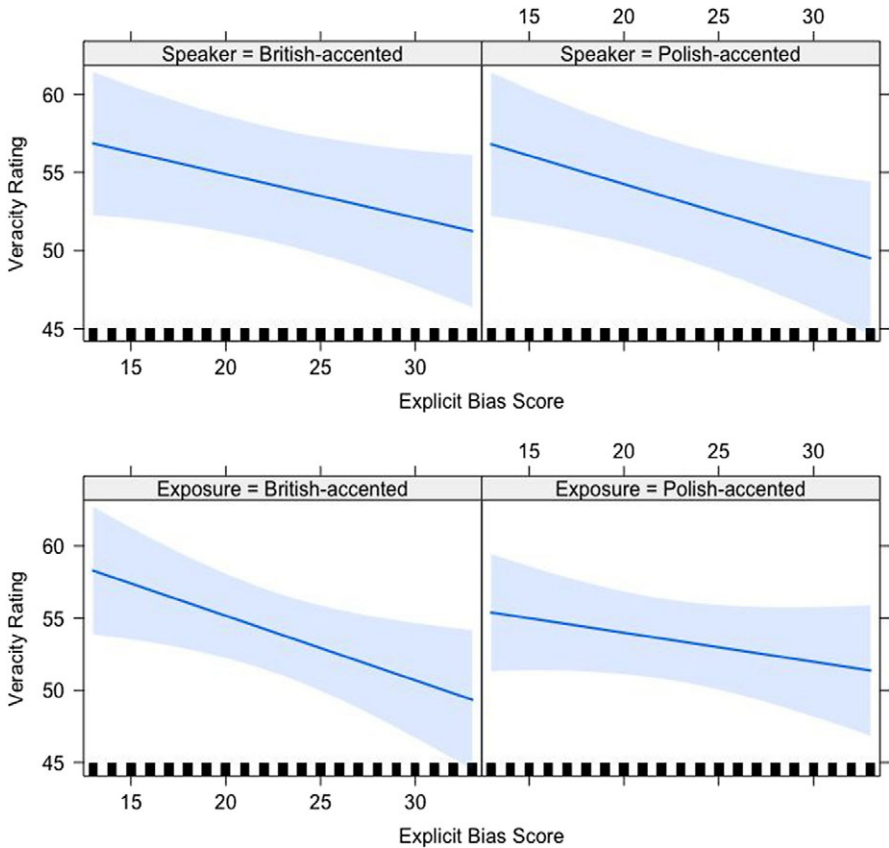


Figure 3. Effect plots of explicit bias on veracity rating.

Note: The top panel shows the effect of explicit bias score on veracity rating by Trivia Speaker; the bottom shows the effect by Exposure. The shaded areas represent 95% confidence intervals.

effects model similar to the initial model above, with the addition of Explicit Bias Score (the composite score) and its interaction terms with the other fixed effects. As with the original model, this model showed no main effect of Trivia Speaker ($\beta = -0.55$, $SE = 2.33$, $t = -0.24$, $p = 0.81$) and no main effect of Exposure ($\beta = 3.05$, $SE = 2.83$, $t = 1.08$, $p = 0.28$). The main effect of Explicit Bias Score was significant ($\beta = -0.32$, $SE = 0.16$, $t = -2.58$, $p = 0.01$). None of the interactions involving the three factors were significant (all p 's > 0.25).

As seen in Figure 3, the main effect of the composite explicit bias score showed that the listeners' veracity judgment scores decreased as their explicit bias scores increased. However, this effect was not modulated by Trivia Speaker (top panel of Figure 3) or Exposure (bottom panel).

Discussion

We conducted a close replication of the study reported by Boduch-Grabka and Lev-Ari (2021), which demonstrated that exposure to Polish-accented English speech modulated subsequent judgments of the veracity of Polish-accented statements and further

demonstrated via a mediation analysis that improved veracity ratings following exposure might be attributable to a decrease in processing difficulty. Hypotheses 1–6, detailed above, summarize the Boduch-Grabka and Lev-Ari (2021) findings; here we summarize our findings with respect to each hypothesis:

- **Hypothesis 1:** We found no evidence that veracity ratings were overall lower for Polish-accented statements than for British-accented statements.
- **Hypothesis 2:** We did not observe an interaction of Exposure condition and Trivia Speaker condition.
- **Hypothesis 3:** Participants in the Polish-accented Exposure condition did not assign higher veracity judgments to Polish-accented statements than did participants in the British-accented Exposure condition.
- **Hypothesis 4:** We observed no effect of Exposure condition on comprehension accuracy for Polish-accented sentences.

To summarize the current findings with respect to Hypotheses 1–4, we did not reproduce the effect of accent on veracity judgments that has been previously reported (Boduch-Grabka & Lev-Ari, 2021; Lev-Ari & Keysar, 2010) or the Boduch-Grabka and Lev-Ari (2021) finding that prior exposure to Polish-accented English speech improves veracity judgments of Polish-accented statements.

- **Hypothesis 5:** Comprehension scores did not predict veracity judgments of Polish-accented statements.
- **Hypothesis 6:** Because we observed no effect of Exposure condition on comprehension accuracy (Hypothesis 4), the conditions for conducting a mediation analysis were not met (Renard, 2019).

We similarly did not reproduce the effects of accent and exposure on comprehension scores (Hypotheses 5 and 6). This finding is consistent with the original authors' argument that processing difficulty operationalized as comprehension scores (partially) explains accent-based veracity judgment effects: if, as we observed here, there is no effect of accent on comprehension scores, no accent-based veracity effect caused by processing difficulty is predicted.

In addition to conducting a close replication of the original study, we elicited participants' self-reported explicit biases relating to "British people" and "Polish migrants". Pre-registered exploratory analyses indicated that the inclusion of a composite measure of explicit bias relating to Polish people significantly improved the mixed effects models of veracity judgments for *both Polish-accented and British-accented statements*, suggesting that explicit bias as measured here did not contribute to an enhanced understanding of listeners' responses to Polish-accented speech in particular. It is also worth noting that a lack of relationship between listeners' explicit biases and their responses to speech accents has been observed in other studies as well (Babel & Russell, 2015; Pantos & Perkins, 2013).

Comparing the current and original studies

Why do the current results differ from those of Boduch-Grabka and Lev-Ari (2021)? It is estimated that approximately one-third to two-thirds of replication studies in the social and psychological sciences have not replicated the original study results

(Camerer *et al.*, 2018; Open Science Collaboration, 2015). Reasons for this lack of replication may be random (e.g., statistical error) or systemic (e.g., bias against publishing null results or research misconduct), and may additionally be attributable to differences in the implementation of the original and replication studies. Methodological discrepancies may result from factors such as insufficient methodological detail provided by the original researchers, original or replication researcher errors, or inherent and unavoidable differences like the time frame when data collection occurred. In Tables 1 and 3 above, we summarized notable methodological comparisons between Boduch-Grabka and Lev-Ari (2021) and the current study.

According to 2021 census data, Polish was the second-most spoken “main language” in England and Wales, unchanged from the 2011 census (Office for National Statistics, 2022). The proportion of short-term residents in England and Wales who were born in Poland decreased between the 2011 and 2021 census (Office for National Statistics, 2023b), while Poland retained its position of the second most frequent country of birth of long-term residents from 2011 to 2021, following a rapid change from 18th to 2nd place between 2001 and 2011 (Office for National Statistics, 2023a). These shifts in migration patterns, in addition to the social and cultural effects of Brexit and the Coronavirus pandemic, may have impacted participants’ performance in unavoidable and unknown ways, such that the date of data collection, as well as the social, cultural, and political milieu should be considered a crucial rather than extraneous variable in future studies.

Participant recruitment also differed between the original and current studies: while the original study’s participants were recruited via contacts, mailing lists, and social media, the current participants were recruited via *Prolific.com*. In addition, the Boduch-Grabka and Lev-Ari’s (2021) participants were “often individuals in public-sector roles” (p. 6), though the authors did not specify the proportion of participants meeting this description. Without this information needed to guide recruitment for the replication study, we made no effort to influence the representation of people holding public sector roles among our participants. Depending on the proportion of the original study’s participants in these roles, and the actual effects of such roles on study performance, this difference between the studies may be responsible for differences in findings. We also do not know how the original study’s participants were compensated and whether possible differences in compensation schemes affected performance on these tasks.

Additional important differences between the participant samples in the two studies may have resulted from differences in exclusionary criteria and their application. For example, the original study’s authors did not detail how screening for “native language” was conducted, and we do not know whether the current screening criteria for language background and locations of birth and residence produced a participant sample with similar language and residential history profiles as the original study. They also did not specify whether participants were screened for Polish language experience; in our more limited data set of 189, we excluded four participants who reported study or knowledge of the Polish language. It is also worth noting that most of our participants reported knowledge of more than one language. Given that the original article did not report the language backgrounds of participants, we do not know whether our sample differed from theirs in this regard; however, participants in the British-accented exposure condition in the current study exhibited much higher comprehension scores than did participants in the same condition in the original study, which could be due to exposure in daily life to various accents. Indeed, exposure to multiple accents has been shown to enhance listeners’ ability to comprehend accented speech (Baese-Berk, Bradlow, & Wright, 2013).

Because we were originally interested in comparing performance by groups of individuals reporting different ethnic social groups, we asked participants to identify the “predominant ethnic composition of their social group, including friends, family, and co-workers” (a pre-registered exploratory analysis was abandoned due to only nine participants indicating a predominantly non-British social group). Our use of the term “predominant” when asking about social groups may have led participants who would have been excluded from the original study (e.g., for having a Polish friend) to qualify for inclusion in the current study. However, we cannot know this because the original authors reported neither the wording of the question used to implement this exclusionary criterion nor the number of participants who were excluded for this reason.

There may also have been important differences in the instructions participants received in the two studies. For example, we did not explicitly instruct participants to use the full range of the veracity scale. It is possible that participants in the original study were given instructions that led them to register a wider range of veracity scores and that this increased the chance of observing differences among the conditions, while our lack of explicit instruction regarding use of the scale led to compressed responses. (However, note that an analysis of z-scored veracity judgments also revealed no differences among the conditions). A number of additional methodological differences not mentioned here may also account for the discrepancies in findings between the two studies.

While we did not replicate the original findings with respect to an accent-based veracity judgment effect or a mitigating effect of accent exposure, it is important to note that the current results should not be interpreted as not replicating the Boduch-Grabka and Lev-Ari (2021) finding that discrepancies between the veracity judgments of Polish-accented and English-accented statements may be attributable to processing difficulty. Indeed, our finding that neither veracity judgments nor comprehension scores differed for British-accented and Polish-accented English speech is consistent with the processing difficulty hypothesis. To the extent that the lack of effect of speaker accent on comprehension scores reflects a lack of difference in processing difficulty for Polish-accented and British-accented speech in this study, the processing difficulty hypothesis does not predict a difference in veracity judgments of speech produced in the two accents.

Limitations and future directions

There are a number of limitations to the original and current studies that are worth considering. One is the relatively small effect size of accent-based veracity effects (as in, e.g., Lev-Ari & Keyser, 2010, as noted by Boduch-Grabka & Lev-Ari, 2021), which may impact the likelihood that replication studies such as the current one would observe such an effect. A related limitation, acknowledged by Boduch-Grabka and Lev-Ari (2021), is that the duration of accent exposure is quite brief, and a longer exposure period might produce a bigger accent exposure effect and thus increase the chances of observing such effects in future studies.

As also acknowledged by Boduch-Grabka and Lev-Ari (2021), this research may be further limited by the presentation of the same audio samples during Polish-accented exposure and the comprehension task. This means that participants in the Polish-accented exposure condition had heard the exact sentence productions during exposure that they were later asked to transcribe for the purpose of assessing comprehension of Polish-accented speech, while participants in the British-accented exposure had not (recall that participants in the British-accented exposure condition had heard the same sentences during exposure, but they were produced by British-accented speakers). This limitation is related to a larger materials-related concern, which is that the voices presented in all the audio tasks came from the same set of 16 speakers (8 British-accented and 8 Polish-

accented English speakers), leading to questions about the generalizability of any effects of exposure. Without examining responses to new speakers, we do not know whether the exposure manipulation here is better characterized as exposure to Polish-accented/British-accented English speech, or as exposure to the speech of these particular groups of speakers. Future studies involving different speakers during the various tasks will help to connect this research to broader populations. Finally, this research was conducted entirely online and remotely, and participants were asked to respond to disembodied voices in a communicative setting that is particular to the research. Ultimately, a fuller understanding of these phenomena will be enhanced by studies of veracity judgments in settings more closely approximating the richness of real-world sociocultural contexts.

Conclusion

Replication studies play an important role in the research cycle, contributing additional data to support an understanding of the robustness of published research studies. We conducted a close replication and extension of the study reported in Boduch-Grabka and Lev-Ari (2021). In contrast to the original study, we did not find that British-English-speaking listeners judged statements produced by Polish-accented English speakers as less likely to be true than statements produced by British-accented English speakers. We also did not find an effect of prior exposure to Polish-accented English speech on veracity judgments, and no effect of exposure on comprehension of Polish-accented sentences. On the basis of these findings, we conclude that the previously observed accent-based reduction in veracity judgments for this populations of listeners, and this particular set of speech materials, may not be robust to replication. Further inquiry into the nature of accent-based veracity judgment effects should consider the replicability of this and similar studies with the goal of determining the factors that influence the presence or absence of accent-based veracity judgment effects as well as the factors that mediate them.

Acknowledgments. This research was supported by a grant from the Council of Dee Fellows, University of Utah, awarded to Rachel Hayes-Harb; an Office of Undergraduate Research (University of Utah) Small Grant awarded to Nathaniel Miller; and Department of Linguistics / College of Humanities research funding awarded to the entire Capstone class. We are additionally grateful to Thomas Bak, Shannon Barrios, Erin Buchanan, Shiri Lev-Ari, Kevin McManus, Dina Mehmedbegovic-Smith, Rafał Jończyk, Seth Wiener, anonymous reviewers, and the study participants for their contributions to this work.

Data availability statement. The experiments in this article earned Open Data and Open Materials badges for transparent practices. The materials and data are available at (<https://osf.io/etgc4/>).

References

- Allen, L., O'Connell, A., & Kiermer, V. (2019). How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRedit) is helping the shift from authorship to contributorship. *Learned Publishing*, 32, 71–74.
- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America*, 137, 2823–2833.
- Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, 133, EL174–EL180.
- Boduch-Grabka, K., & Lev-Ari, S. (2021). Exposing individuals to foreign accent increases their trust in what nonnative speakers say. *Cognitive Science*, 45, e13064.
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmeld, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., ... Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2, 637–644.

- Castillo, P. A., Tyson, G., & Mallard, D. (2014). An investigation of accuracy and bias in cross-cultural lie detection. *Applied Psychology in Criminal Justice*, 10, 66–82.
- Cheng, L. S. P., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The problematic concept of native speaker in psycholinguistics: Replacing vague and harmful terminology with inclusive and accurate measures. *Frontiers in Psychology*, 12, 715843.
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116, 3647–3658.
- Dunin-Wasowicz, R. D. R. M. (n.d.). Incomplete Europeans: Polish migrants' experience of discrimination in the UK is complicated by their whiteness. *London School of Economics Brexit 2015–2021 Blog*. <https://blogs.lse.ac.uk/brexit/2016/03/09/incomplete-europeans-polish-migrants-experience-of-prejudice-and-discrimination-in-the-uk-is-complicated-by-their-whiteness/> (Original work published 2016)
- Hanzlíková, D., & Skarnitzl, R. (2017). Credibility of native and non-native speakers of English revisited: Do non-native listeners feel the same? *Research on Language and Computation*, 15, 285–298.
- Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46, 1093–1096.
- Liu, M., Chong, S. W., Marsden, E., McManus, K., Morgan-Short, K., Al-Hoorie, A. H., Plonsky, L., Bolibaugh, C., Hiver, P., Winke, P., Huensch, A., & Hui, B. (2023). Open scholarship in applied linguistics: What, why, and how. *Language Teaching*, 56, 432–437.
- Office for National Statistics. (2022). *Language: England and Wales, Census 2021*. <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language/bulletins/languageenglandandwales/census2021>
- Office for National Statistics. (2023a). *The changing picture of long-term international migration, England and Wales: Census 2021*. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/thechangingpictureoflongterminternationalmigrationenglandandwales/census2021>
- Office for National Statistics. (2023b). *The changing picture of short-term international migration, England and Wales: Census 2021*. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/thechangingpictureofshortterminternationalmigrationenglandandwales/census2021#short-term-residents-by-country-of-birth-and-passport-held-and-how-this-has-changed-over-time>
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716.
- Pantos, A. J., & Perkins, A. W. (2013). Measuring implicit and explicit attitudes toward foreign accented speech. *Journal of Language and Social Psychology*, 32, 3–20.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51, 195–203.
- Portas, A. (2018). Representations of Polish migrants in British media from the perspective of “Moral Panic” Theory. *Polish Journal of English Studies*, 4, 113–142.
- Porte, G., & Mcmanus, K. (2019). *Doing Replication Research in Applied Linguistics*. Routledge.
- Renard, M. (2019). Doing and reporting your first mediation analysis in R. *Towards Data Science*. <https://towardsdatascience.com/doing-and-reporting-your-first-mediation-analysis-in-r-2fe423b92171>
- Rzepnikowska, A. (2019). Racism and xenophobia experienced by Polish migrants in the UK before and after Brexit vote. *Journal of Ethnic and Migration Studies*, 45, 61–77.
- Souza, A. L., & Markman, A. B. (2013). Foreign accent does not influence cognitive judgments. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 35, 1360–1365.
- Stocker, L. (2017). The impact of foreign accent on credibility: An analysis of cognitive statement ratings in a Swiss context. *Journal of Psycholinguistic Research*, 46, 617–628.
- Wetzel, M., & Gygax, P. (2021). Do non-native and unfamiliar accents sound less credible? An examination of the processing fluency hypothesis. *Journal of Articles in Support of the Null Hypothesis*, 17, 1539–8714.

Cite this article: Barlow, S., Beardsley, G., Bsharah, Z., Crofts, R., De La Rosa, C., Gutierrez, A., Highfill, C., Johnson, A. G. W., Johnson, C., Johnson, J., Cardenas, I. L., Martinez, J. T., Miller, N. T., Murray, R. M., Page, S., Petersen, T., Ramos, I., Rhodes, R., Vainuku, P., Wednesday, B. M., Farnsworth, E. C., Kim, S. K., & Hayes-Harb, R. (2024). The effects of exposure and explicit stereotypes on veracity judgments of Polish-accented English speech: A preregistered close replication and extension of Boduch-Grabka & Lev-Ari (2021). *Studies in Second Language Acquisition*, 1–17. <https://doi.org/10.1017/S0272263124000123>