Dynamic sociolinguistic processing: Real-time changes in judgments of speaker competence

Erez Levon, Devyani Sharma, Yang Ye

Language, Volume 98, Number 4, December 2022, pp. 749-774 (Article)

Published by Linguistic Society of America
DOI: *https://doi.org/10.1353/lan.2022.0020*

DYNAMIC SOCIOLINGUISTIC PROCESSING: REAL-TIME CHANGES IN JUDGMENTS OF SPEAKER COMPETENCE

Erez Levon
University of Bern

Devyani Sharma
Queen Mary University of London

Yang Ye
University of Greenwich

Social category information plays a central role in speech perception and processing. To date, research on this topic has struggled to model how social category perceptions evolve over the time course of an interaction. In this article, we build on recent methodological developments to investigate trajectories of listener perceptions, focusing on how impressions change as linguistic, social, and contextual details emerge. We base our arguments on an analysis of listeners' real-time evaluations of the perceived competence of speakers of two British regional accents during a mock interview for a job in a law firm. Results indicate a need to move away from the view, predominant in sociolinguistics, of category perception as a discrete phenomenon and toward a model of perception as inference under uncertainty. We discuss implications for theories of sociolinguistic cognition and for understandings of accent bias.*

Keywords: sociolinguistic perception, real-time methods, attitudes, accent bias, dynamic construal

1. Introduction. Associating linguistic forms with social categories is central to speech perception and processing. Over the past twenty years, research has demonstrated that these associations play a crucial role in phenomena as diverse as phoneme discrimination (e.g. Johnson, Strand, & D'Imperio 1999, D'Onofrio 2015), lexical access (e.g. Walker & Hay 2011, Kim & Drager 2018), language comprehension (e.g. Babel & Russell 2015, Lev-Ari, Ho, & Keysar 2018), and language change (e.g. Harrington et al. 2018), not to mention speaker evaluation and impression formation (e.g. Campbell-Kibler 2008, Preston 2010, Levon 2014). Insights from these studies have allowed us to develop sophisticated models of speech processing and perception, and the various social, psychological, and cognitive factors that constrain it.

Yet to date, such studies have tended to abstract away from the inherently temporal dimension of language—the fact that social category perceptions unfold in real time with cumulative exposure to linguistic input. Research has instead tended to focus on how perceptual outcomes affect downstream processing. Niedzielski (1999), for example, demonstrates that listeners' classifications of different variants of the onset of the MOUTH diphthong in English depend on whether listeners were primed to believe the speaker was Canadian or American (see also Hay & Drager 2010). Similarly, Pharao et al. (2014) show that, among the listeners they tested, evaluative judgments of fronted [s] in Danish depended on listeners' beliefs about the speaker. In both studies, processing and evaluation depend on a prior judgment of social category membership.

The role of context in social category perception has been noted in recent work. Hilton and Jeong (2019), for instance, show that the common or stereotypical meanings of certain linguistic variants can be obscured by contextual enrichment, rendering features indexically inoperative in certain contexts (see also Campbell-Kibler 2009 on so called 'bullet-proofing' and Dragojevic & Giles 2014 on reference frames). Levon and Ye (2020) similarly argue that uptalk in Standard Southern British English takes on different meanings depending on whether it is used in a (mock) rape trial or a (mock) med-

* We gratefully acknowledge the support of the UK Economic and Social Research Council (grant ref: ES/P007767/1) for the research reported here. Special thanks also to Kathryn Campbell-Kibler, Lauren Hall-Lew, Andries Coetzee, and two anonymous referees for their detailed comments on earlier versions of our arguments. We alone are responsible for any remaining errors or shortcomings.

ical malpractice trial. These studies demonstrate that perceived meanings do not exist in a vacuum, but rather arise in specific situated contexts (Eckert 2012). But in this more recent work too, the temporal dimension and the fact that contextual information itself unfolds over time has not been the focus.

With some notable exceptions, studies of social processing effects have mostly relied on fixed situational contrasts and RETROSPECTIVE evaluations, that is, endpoint evaluations offered after an impression of a speaker and/or context has been formed. However, linguistic perception is inherently temporal: we encounter language in REAL TIME, and our classifications and evaluations of speakers evolve as we are exposed to additional information from the ongoing speech stream. As both accent and content/context unfold, participants make fast online decisions about whether to integrate new information and update their prior beliefs about the situation. Understanding how listeners dynamically update their evaluations over the course of an interaction is a crucial, and currently largely missing, component of a comprehensive model of sociolinguistic perception. How much do sociolinguistic processing and judgments change in real time? Does this differ across individuals? In short, what is the process that leads to the many outcomes (evaluations) we have seen in the literature to date?

In this article, we build on recent methodological developments in tracking listener evaluations in real time (Watson & Clark 2013, 2015, Hesson & Shellgren 2015, Montgomery & Moore 2018, Levon, Buchstaller, & Mearns 2020, Austen & Campbell-Kibler 2022) to investigate the trajectory of the listener evaluation process. We focus on how impressions change over the course of an utterance, as linguistic, social, and contextual detail emerges, and whether these change in the same way for different listeners and talkers. To do so, we move away from a static and discrete view of linguistic variants as either signaling category membership or not. Instead, we ground our discussion in a Bayesian view of perception as inference under uncertainty (Frank & Goodman 2012, Kleinschmidt & Jaeger 2015, Kleinschmidt, Weatherholtz, & Jaeger 2018, Burnett 2019), whereby listeners link variants to categories probabilistically. As the speech stream continues and new information is encountered, probabilities may be adjusted and listener impressions evolve. Only by examining evaluations from this time-linked perspective, we argue, can we gain insight into the process of impression formation itself.

We begin with a brief overview of current models of sociolinguistic perception and evaluation, focusing primarily on Niedzielski and Preston's (2003) procedural model of LANGUAGE REGARD. We then describe how this model can be integrated with a dynamic, probabilistic approach to perception and category activation taken in cognitive science (Kleinschmidt & Jaeger 2015). We go on to present an analysis of listeners' real-time evaluations of the perceived competence of speakers of two British regional accents during a (mock) interview for a job in a law firm. Using a series of GENERALIZED ADDITIVE MIXED MODELS (GAMMs; Wood 2017) of listener evaluations across the full time course of a speaker's utterance, we show that listeners' initial evaluations are affected in different ways by unfolding content and context, which shapes the TRAJECTORY of impression formation. Some evaluation trajectories show early updating, while others show late or minimal change, and these differences relate to listener age and talker accent. We discuss implications for current models of sociolinguistic cognition such as language regard. Drawing on theories of dynamic person construal (Freeman et al. 2008, Freeman et al. 2010) in psychology, we also propose TRAJECTORY BIAS as an important component of accent bias.

2. PATHWAYS OF SOCIOLINGUISTIC PERCEPTION.

The study of listener attitudes and impressions has been central to the sociolinguistic enterprise since the inception of the field (Labov 1972, Kristiansen 2010). Various models have been proposed for the underlying sociocognitive process leading to these evaluative outcomes. The most established within sociolinguistics is Niedzielski and Preston's (2003) procedural model of language regard.

For Niedzielski and Preston, the first step of sociolinguistic perception is 'noticing', a process of form-based discrimination. Noticing requires a listener to recognize (at whatever level of conscious awareness) that a given linguistic output is one option among several, thus setting up a contrast between the realized form and alternatives. Noticing is sensitive to a variety of factors, often described under the rubric of salience (Kerswill & Williams 2002, Drager & Kirtley 2016). These include properties of linguistic variables themselves (such as their psychoacoustic prominence or their semantic transparency; e.g. Silverstein 1981, Yaeger-Dror 1993, Preston 1996) and the conditions under which a reaction has been elicited (Kristiansen 2009, McKenzie & Carrie 2018). Once a linguistic form has been noticed, Niedzielski and Preston (2003) argue that it is then 'classified', that is, associated with a given social group or speech context. Classification relies on the prior existence of a cultural model (Gal 2016), stored in memory, that links particular ways of speaking with particular speakers and situations. Like noticing, classifying is also subject to a range of external factors, including the social and/or linguistic context (Campbell-Kibler 2011, Levon 2014, Pharao et al. 2014), the cumulative distribution of variants across an utterance (Labov et al. 2011, Wagner & Hesson 2014, Levon & Buchstaller 2015), and a listener's prior experience with a relevant feature (Walker & Hay 2011, Hay, Drager, & Gibson 2018). Via a process of activation spreading, classification of a form as belonging to a particular social category leads to the third step, namely 'imbuing' a form with various affective and evaluative meanings associated with that category (Macrae & Bodenhausen 2001). These imbued meanings give rise to observed behavioral reactions to language, the fourth and final step, which includes both reactions that involve more cognitive deliberation (e.g. subjective evaluations; Cargile et al. 1994, Cargile & Giles 1997) and those that do not (e.g. phonetic imitation in a lexical shadowing task; Babel 2012, and see also Pantos 2019). Over time, the link between classifying and imbuing can become routinized, leading to the formation of automatic associations, stored in memory, between linguistic forms and evaluative meanings (Preston 2019).

This nuanced model was developed over the last twenty years using retrospective (endpoint) evaluations. Recent work has begun to explore listener reactions across the time course of an utterance. A real-time approach to the study of sociolinguistic perception was first adopted by Watson and Clark (2013, 2015). Using a bespoke real-time reaction tool, Watson and Clark (2013) considered the extent to which judgments of the perceived status of speakers from two locations in northwest England were affected by merged tokens of NURSE and SQUARE at the moment of encountering them. Listeners' sensitivity was dynamic, constrained not only by general properties of the variants (whether they coincide with standard British pronunciations) but also by the unfolding linguistic context (whether the nonstandard pronunciation was encountered before or after a standard token). Watson and Clark (2015) largely replicate this finding with four further varieties. Levon, Buchstaller, and Mearns (2020) apply Watson and Clark's (2013, 2015) real-time reaction tool to examine the cumulative real-time effect of phonetic and morphosyntactic features of Newcastle English on the perceived status of a speaker, and they provide further support for the idea that impression formation is dynamic and evolves over the course of an utterance (see also Hesson & Shellgren 2015).

The above studies focus on the final step in the language-regard process, the evaluative reaction. Montgomery and Moore (2018) examine the earlier steps of noticing and classifying. Using a new reaction tool, they asked respondents to listen to speakers from the Isles of Scilly, off the southwest coast of England, and to indicate, while listening, any time they encountered a feature that might give information about where the speaker was from. Montgomery and Moore found that the content of the extract and other co-present phonetic variants influenced whether features were noticed and how they were classified. With a more global evaluation task, Austen and Campbell-Kibler (2022) found that continuous evaluation did not closely track specific cues, and that in-the-moment and after-the-fact ratings may engage different cognitive processes.

The findings of these prior studies resonate with recent proposals in social psychology to adopt a less mechanistic understanding of how language attitudes emerge, and to conceptualize the evaluation process as one of active sense-making within a situated social context (Giles & Marlow 2011, Giles & Rakić 2014). Giles (2011) describes such sense-making as a product of MOTIVATED INFORMATION MANAGEMENT (Afifi & Weiner 2004), whereby a listener becomes aware of the difference between their current level of uncertainty regarding a particular topic (e.g. their evaluation of a speaker) and their desired level of uncertainty. This 'uncertainty gap' leads listeners on a targeted search for information that they believe—given the current interactional context, as well as their own social histories and experiences—will allow them to achieve their desired knowledge state. As additional information is acquired, a listener's uncertainty gap shifts, leading them to reassess what information it is useful for them to attend to in the speech signal.

Giles's (2011) dynamic characterization of perception as the management of uncertainty in real time provides an elegant conceptual framework for capturing the kinds of contextual variability found in the recent sociolinguistic perception work reviewed above. The model of language regard complements this by identifying the stages of linguistic perception and the subtle factors that facilitate or inhibit activation of social representations. However, the model has taken a relatively static binary focus for classifying category membership, namely whether or not a category has been activated.

In order to capture the reality of uncertainty in real time, we propose replacing the idea of activation as discrete and categorical (on or off) with a probabilistic model. In a formulation reminiscent of the 'uncertainty gap', Kleinschmidt and Jaeger (2015) describe the perception of both social and linguistic information as a process of INFERENCE UNDER UNCERTAINTY (see also Kleinschmidt, Weatherholtz, & Jaeger 2018; Burnett 2019 takes a similar approach). We follow Kleinschmidt and Jaeger (2015) here, but several recent psycholinguistic models have proposed similar dynamic interactive systems (see Getz & Toscano 2021 for a review). Similarly for category activation, Casasanto and Lupyan (2015) take account of dynamic, contextual, and experiential timescales and offer a review of earlier work that emphasizes this dynamic perspective.

Rather than treating linguistic forms as simply activating categories, Kleinschmidt et al. argue that they activate conditional probability distributions, that is, inferential calculations of the likelihood that a given social meaning should be inferred by a particular form in a specific context. Crucially, probabilistic models treat these inferences as subject to continuous updating as new information is encountered.

The suggestion by Kleinschmidt and others to treat perception as a process of dynamic competition between multiple, partially activated meanings is an important first step toward a more socially and psychologically realistic model of the role of language in social impression formation. Focusing on competition among meaning potentials is also in keeping with recent advances in work on sociolinguistic production, which has argued forcefully that linguistic forms are linked to a variety of potential meanings rather than to an exclusive one (e.g. Eckert 2012), as well as with current exemplar-theoretical accounts in phonology and psycholinguistics, in which competition among multiple possible underlying representations is crucial (Johnson 2006, Sumner et al. 2014, Hay 2018).

In the present study, we uncover this incremental updating of initial probabilities ('priors', in Bayesian terms) in real time, as context unfolds. In line with recent research reviewed above, the findings demonstrate differential degrees of influence exerted by the unfolding context on the emergent social interpretation of linguistic forms and on perceptual outcomes for different listeners. We examine listeners' real-time evaluations of speakers of two socially differentiated British accents (Received Pronunciation and Multicultural London English) in a particular situational context (a job interview in a law firm). Doing so allows us to consider how both long-standing hierarchies of accent prestige in Britain—priors, in the minds of listeners—and immediate situational expectations and content affect the impression-formation process. Unlike most previous work on real-time judgments of speech (though cf. Hesson & Shellgren 2015), we do not concentrate only on listeners' reactions to particular accent features. We consider the full trajectory of perceptual responses over time for a holistic picture of the social and cognitive processes underlying perception and speaker evaluation.

3. METHOD.

3.1. BACKGROUND. This study is part of a larger project (Levon, Sharma, & Watt 2017–2021) investigating current attitudes toward accent variation in the UK and the extent to which these attitudes may impede professional advancement for certain social groups. Social mobility is widely considered to be stagnant in the UK (Social Mobility Commission 2019). Despite repeated government commitments to tackle the issue (Cabinet Office 2011, Mason 2013), research has demonstrated that upward income mobility in Britain has declined over the past fifty years (e.g. Blanden, Gregg, & Machin 2005), that elite professions like law and medicine continue to be dominated by people from socially and economically privileged backgrounds (Friedman, Laurison, & Miles 2015), and that the social status of one's family remains the strongest predictor of attained levels of wealth, education, and asset ownership in the UK (Clark 2014).

It is well known that, particularly in the UK, accent encodes many of the social contrasts (e.g. ethnicity, parental social class, region of origin) that underpin these differences in attainment. For centuries, accent has played a constitutive role in signaling class and education in British society (Shaw 1916, Sheridan 1762, Mugglestone 2007), fueling discriminatory stereotypes about speakers of nonstandard varieties. Accent is a very likely contributor to observed patterns of employment bias and unequal access in Britain.

The Accent Bias in Britain project adopts a multimethod approach to examining attitudes toward accents and their effect on social mobility. Project research to date has included a nationwide survey of the UK public's explicit evaluations of accent labels (Sharma, Levon, & Ye 2021), verbal-guise studies of judgments of professional competence of speakers of five accents in England among both expert and nonexpert listeners (Cardoso et al. 2019, Levon et al. 2021), and experiments assessing the effectiveness of anti-bias interventions in reducing accent-based disparities (Ye et al. 2020). The current study adds the dimension of listeners' perceptual reactions in real time. We aim to address three primary questions:

(i) How do attitudes toward language and evaluative impressions of speakers form during the time course of listening, and do these evaluation trajectories differ across individuals?

(ii) How do trajectories of listeners' dynamic evaluations compare to more commonly used retrospective ratings?
(iii) What does this kind of real-time analysis add to current understandings of perception and evaluation as a social and psychological process, and how does this impact the ways we approach accent bias and discrimination in society?

3.2. DESIGN. We conducted a verbal-guise study (Cooper & Fishman 1974) with a graphical sliding scale to capture listeners' real-time evaluations of speakers of two English accents: Received Pronunciation (RP) and Multicultural London English (MLE). RP is the national standard accent in England, historically associated with prestige and authority (Fabricius 2005, Mugglestone 2007). While attitudes toward RP may have shifted somewhat over the past three decades, studies have shown that RP remains the most overtly prestigious accent in the British language ideological landscape (Giles 1970, Coupland & Bishop 2007, Sharma et al. 2021). MLE, in contrast, emerged more recently, over the last twenty years, within multiethnic, working-class adolescent friendship groups in Inner London (Cheshire et al. 2008, 2011, Fox 2015). Though MLE carries local prestige, particularly among young people, there is evidence of wider negative bias against MLE speakers (Kircher & Fox 2019, Levon et al. 2021). We therefore chose to compare listeners' real-time reactions to RP versus MLE because they sit at opposite ends of the prestige hierarchy in Britain today. (A list of phonetic features observed for each accent in the recordings is provided in Appendix A.)

MATERIALS. Stimuli for the present study were drawn from the pool of stimuli developed as part of Levon et al. 2017–2021. For the larger project, we recorded ten young men (aged eighteen to twenty-five) who were native speakers of five different regional accents in England (two speakers per accent). The stimuli consisted of fifteen short (c. thirty-second) mock responses to questions asked in an interview for a junior lawyer position in a law firm. Our focus on law firm interviews provides listeners with a specific social context in which norms of prestige, knowledge, and competence are relevant. The content of the responses was developed in close consultation with lawyers on the project's advisory board, and covered a range of typical interview topics, some requiring technical legal expertise and some about more general professional skills. Written versions of all question–response pairs were pretested on a group of twenty lawyers otherwise unrelated to the project, who confirmed their naturalness and relative quality. We aimed for response quality that would be judged as neither very good nor very bad, since research in psychology has demonstrated that peripheral cues (like accent) play a more active role in determining evaluative judgments when the content to be evaluated is ambiguous (e.g. Chaiken & Maheswaran 1994).

Recorded stimuli were post-processed to remove disfluencies or hesitation markers, to standardize intensity (at 70 dB), and to ensure that mean pitch levels of all recordings were comparable. We did not control for speaking rates across speakers, under the assumption that speech rate may itself be a dialect feature (e.g. Jacewicz, Fox, & Salmons 2007). Recordings were pilot-tested with (i) ten linguists with expertise in British dialect variation to confirm that the recordings provided accurate representations of the target accents, (ii) sixty-eight listeners from the UK general public to confirm that speakers for a given accent were evaluated similarly in terms of traits like perceived 'masculinity' and 'friendliness', and (iii) an additional 130 listeners from the UK general public to confirm that the recordings did not sound 'fake' or 'forced'.

For the present real-time study, we selected four pairs of mock answers, one set produced by a speaker of RP and the other by a speaker of MLE (see Appendix A). Two questions required expert legal knowledge and two were about more generic professional skills, allowing an examination of the effect of expert content while controlling for question-specific effects. We recognize that using one speaker per accent potentially limits the generalizability of our findings, but rely on the cross-question comparison to provide a kind of sample-internal replication. The stimuli were selected from the larger set of question–response stimuli based on their similar density of MLE features and similar retrospective ratings within accent (Levon et al. 2021).

PARTICIPANTS. Stimuli were presented to 160 UK-based listeners via an online survey (built in Qualtrics). Listeners, who all reported speaking English as a native language, were recruited via Prolific, a web-based platform for locating participants from the general public. Of the 160 respondents, fifty-seven identified as men, 102 identified as women, and one identified as neither a man nor a woman. Respondents ranged in age from eighteen to seventy-three, with a median age of thirty-seven. We did not collect information on respondents' ethnicity or social-class background as these parameters were not shown to constrain attitudinal responses in our previous studies (e.g. Levon et al. 2021).

PROCEDURE. In the experiment, listeners were told that they would be providing two types of evaluations of candidates for a trainee solicitor position at a commercial law firm in the UK. The first type of evaluation was a continuous rating of the candidate using a graphical sliding scale on the computer screen while the candidate was responding to the question (see Figure 1). The sliding scale, which was built specifically for the project and integrated into the Qualtrics survey, was anchored at 0/Doing poorly, on the left, and 100/Doing well, on the right. The slider was positioned at 50 (neutral) at the start of each recording, and respondents were instructed that they could move the slider whenever their judgment of the candidate's likelihood of doing well in the interview changed. Rating movement typically started two seconds after the start of the stimulus. Respondents received immediate numeric feedback directly above the scale as they moved the slider right and/or left, so they were always aware of the slider's numeric position on the scale. The second type of evaluation was an additional retrospective (post-audio) judgment of how well the candidate had done overall and the extent to which the candidate's response had displayed expert knowledge.

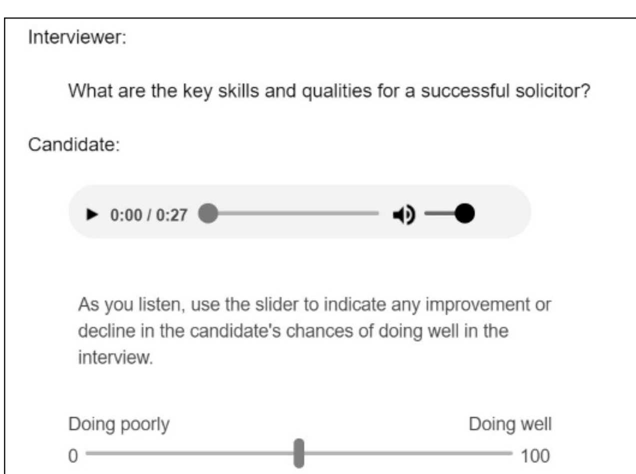

FIGURE 1. Layout of the graphical sliding scale for real-time evaluations.

Following a practice trial (using a North American voice) to familiarize them with the rating instruments, respondents listened to and evaluated four experimental stimuli, two from each accent (RP, MLE). For each accent, respondents heard one response requiring technical expertise ('expert' response) and one response about more general professional skills ('nonexpert' response). In this way, respondents heard and evaluated both an expert and a nonexpert response per accent, never hearing the same response twice. The order of response types (expert, nonexpert) and speaker/accent (RP, MLE) was counterbalanced across the respondent population. After having heard and evaluated all four stimuli, respondents provided basic demographic information. They also completed a short questionnaire about how important it is to them to appear nonprejudicial in their interactions with others. The questionnaire was designed to measure respondents' MOTIVATION TO CONTROL A PREJUDICED RESPONSE (MCPR; Dunton & Fazio 1997), an individual difference measure of the extent to which respondents are driven to engage in more deliberative, as opposed to more automatic, processing as a way of avoiding potentially prejudiced reactions. Our prior work using retrospective evaluation methods showed MCPR to significantly constrain attitudinal outcomes (Levon et al. 2021), so we wished to test for any such effects on trajectories of attitude formation.

MODELING. Respondents' real-time evaluations were examined via a series of GAMMs (Wood 2017; see also Sóskuthy 2017) using the mgcv package (Wood 2011) in R version 4.0.2 (R Core Team 2020). GAMM modeling was chosen as it allows us to examine the entire trajectory of evaluations across time without assuming an a priori shape of the response curves, and to consider the effect of both fixed and random factors on the trajectories observed. The graphical sliding scale that we used automatically recorded every change in rating with an increment of at least one (on a graphical scale of 1–100) and associated that change with a time stamp (sensitive to 1 ms). For the purposes of GAMM modeling, scores were converted to standardized polls of the slider position at 100 ms intervals. For example, if a respondent moved the graphical slider at 150 ms from 50 (the starting point) to 75, this was recorded as a rating of 50 at 100 ms and 75 at 200 ms. This resulted in approximately 300 ratings per respondent for each of the thirty-second stimuli. We subsequently collapsed these fine-grained response intervals into one-second intervals (i.e. the position of the slider at 1 s, 2 s, etc.) in order to allow for easier model computation and to reflect the likely granularity involved in moving a graphical sliding scale after encountering a linguistic feature.

Prior to modeling, real-time measures were converted to relative time (percent time of the stimulus) in order to allow for comparisons of response trajectories across speakers, who had slightly different speech rates. Separate models were built for each of the four question–response pairs to allow us to examine the effect of stimulus content on evaluation trajectories. All models were manually stepped down from maximal models, which included parametric terms (fixed effects) for accent and order of presentation; smooth terms (curves) for relative time, respondent age (mean-centered and standardized), and respondent MCPR (mean-centered and standardized); difference smooths for relative time, respondent age, and respondent MCPR by accent (to determine whether the curves associated with these factors vary across the two accents); and all possible tensor product interactions (analogous to interaction terms in standard regression models).[1] All models also included either random smooths for respondents or an error model

[1] Analyses revealed that order (of accent and of question type) had a significant parametric (fixed) effect on evaluations. No interactions were identified between order and time, indicating that order of presentation does not significantly impact response trajectories. In the interest of space, we therefore do not further consider the parametric order effects.

to adjust model outputs for residual autocorrelation, depending on which strategy yielded better results (see Sóskuthy 2017, Baayen et al. 2018). Model smooth terms were built using thin-plate regression splines, and models were fitted using maximum likelihood estimation. Model comparison was done using the compareML function in the itsadug package (van Rij et al. 2020). Full model details are listed in Appendix B.[2]

3.3. PREDICTIONS. Research on accent perception has found very diverse kinds of listener reactions, so we consider three hypothetical types of evaluation trajectories, across a range of sensitivity to accent:

- TYPE 1—ABSENCE OF BIAS: There is no influence of accent (RP vs. MLE) on listeners' perceptions of competence in a job interview. A type 1 response pattern would show no difference in ratings of MLE and RP over the time course of the utterance. Any fluctuations in ratings would correspond to other factors (e.g. listener, content of the response), so would be shared by MLE and RP.
- TYPE 2—EARLY BROAD CATEGORIZATION: Listeners categorize the speaker early based on general recognition of an accent. This early categorization may be subject to updating across the utterance. A type 2 response pattern would show early divergence of MLE and RP in the time course of the reaction.
- TYPE 3—ACCENT-SPECIFIC SHIBBOLETHS: Listener evaluations are sensitive to specific accent features or clusters of features within an utterance. A type 3 response pattern would be perturbed by specific accent features, with RP and MLE fluctuating differently.

These three predicted types of listener response patterns represent a continuum, rather than discrete categories, and could coexist in longer interactions. In the remainder of our discussion, we examine which of these is most apparent in the present data set.

4. RESULTS. Each of the four stimuli (question–response pairs) was modeled individually in order to observe the role of unfolding content (e.g. expert knowledge) and accent features in influencing evaluative responses. The analysis revealed that evaluation trajectories of the two 'expert' stimuli resembled one another, as did the two 'nonexpert' stimuli, and the two types differed from one another. We therefore use findings from one question each to illustrate the overall patterns observed; see Appendix C for parallel results for the other two stimuli. We present reactions first to nonexpert stimuli, followed by expert stimuli.

4.1. EVALUATIONS OF NONEXPERT STIMULI. Listeners exhibited significant differences in their response trajectories to nonexpert stimuli produced in RP and in MLE (see Appendix B for model details). These differences are shown in the top panel of Figure 2, which presents model predictions of evaluations of responses to the nonexpert question 'Why do you want to be a lawyer?'. Gray error bands show the 95% confidence interval across the evaluations of all 160 listeners.

The top panel of Fig. 2 shows an upward trajectory in ratings of both the MLE and the RP speaker over the course of the utterance, indicating an increasingly favorable reaction. However, the slope of the trajectory is significantly steeper for RP than for MLE. After having heard approximately 30% of the utterance (two of six sentences), listeners begin to rate the RP speaker as doing significantly better despite both speakers uttering identical responses ($\chi^2 = 278.42$, $df = 16$, $p < 0.000$). Once it emerges, the pat-

[2] Stimulus recordings, rating scales, dialect density measurements, and rating response data are available via the UK Data Service at https://reshare.ukdataservice.ac.uk/854405. Code for the graphical sliding scale is available at https://github.com/Cardoso-Dialects/AccentBiasBritain.

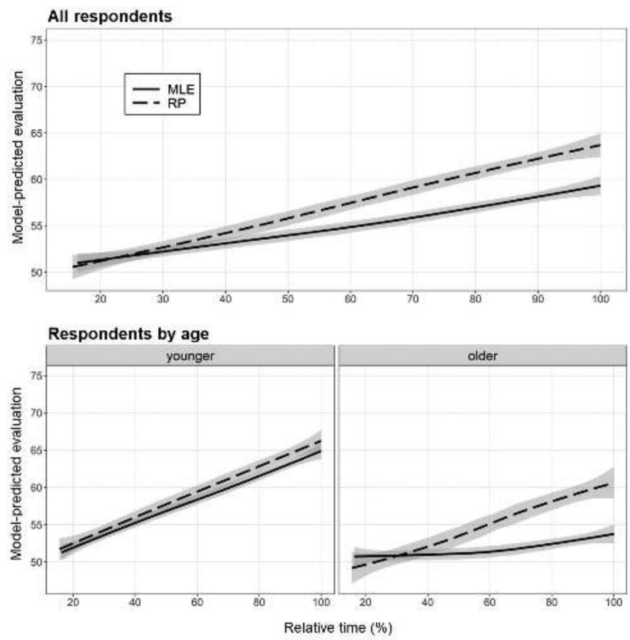

FIGURE 2. Model-predicted evaluations of responses to a nonexpert question.

tern shows dynamic stability. From the onset of differentiation, the RP slope is approximately 70 percent steeper than the MLE slope. Not only do listeners arrive at significantly different endpoints in their ratings of the RP and MLE speakers, but for the final 70 percent of the response listeners also appear to have already decided that the RP speaker was doing significantly better than the MLE speaker. This points to accent bias within the trajectory of the response, in a direction consistent with dominant attitudes toward RP and MLE in the UK (as observed with multiple talkers: Levon et al. 2021, and in explicit attitudes: Kircher & Fox 2019).

The difference is also linear: GAMM modeling does not assume a linear response pattern for the dependent variable, but the model-predicted values depicted in the top panel of Fig. 2 (plotted using locally estimated scatterplot smoothing) appear to be linear in nature. Model results confirm that the smooths associated with the trajectories of evaluations for the MLE and the RP speaker both have EFFECTIVE DEGREES OF FREEDOM (*EDF*) values of 1.000 (the closer the *EDF* value is to 1, the more linear the trajectory).[3] This

[3] In standard linear regression, model degrees of freedom reflect the total number of (linear) predictors in a model. In GAMMs, in contrast, (nonlinear) predictors are computationally smoothed to provide a best estimate of response trajectory, within stated model parameters. The *EDF* value represents a model-generated estimate of the number of distinct linear interpolants necessary to approximate the observed response trajectory. If only one interpolant is required ($EDF = 1$), the trajectory is effectively linear. The more interpolants required (and hence the higher the *EDF* value), the 'wigglier' the response trajectory. For details, see Wood 2017.

Models were built using treatment contrasts, with MLE as the reference level for accent. This means that the model smooth term for the effect of time overall reflects the response trajectory for MLE, that is, the reference level ($EDF = 1.000$, $F = 26.28$, $p < 0.000$). The *p*-value for this term indicates that the trajectory is significantly different from zero (i.e. there is an effect of time). The difference smooth for time by accent, in contrast, reflects the response trajectory for RP ($EDF = 1.000$, $F = 4.20$, $p = 0.04$). Here, the *p*-value indicates that the trajectory for RP is significantly different from the trajectory for MLE. The significance of accent is further confirmed by the model comparison statistics provided.

means that whatever factor is responsible for the upward trajectory observed, its effect is consistent over the course of the response. We later compare this to a different trajectory shape with expert stimuli. This result suggests that listeners bring prior accent-based expectations to the listening event and quickly categorize speakers according to these, giving rise to distinct trajectory slopes for RP (steeper/more positive) versus MLE (flatter/less positive).

Which of the three predicted types of response does the top panel of Fig. 2 resemble? Type 1 (no accent effect) is ruled out by the evident difference in how the two accents are rated. Distinguishing between type 2 (early broad categorization) and type 3 (shibboleth tracking) is more subtle. A sufficient number of distinctive phonetic features arise within the first few seconds of the recordings for the two accents to be identifiably different (eighteen within the first 30% of the recordings; see Appendix A), so the early categorization we see (type 2) does involve some sensitivity to phonetic cues, that is, an element of a type 3 response. However, another forty-six MLE/RP features later in the recordings do not prompt significant changes. In fact, all of the distributions we present conform to this pattern of early categorization (type 2), with modest later adjustments that are not linked to specific accent shibboleths. Rather than tracking accent detail (type 3), this suggests rapid early categorization of speakers (type 2).

This may partly be an artefact of the task and stimuli in question. A salient taboo form, such as swearing or vernacular lexicon, might have triggered more localized reactions of type 3. Indeed, Figure A1 in Appendix C shows that one response set evidenced a marked downward turn after a mention of debt as a career motivation. Further work will help to further specify constraints on dynamic trajectories. Even if the tendency for a type 2 response is partly due to the task at hand, we will see that systematic divergences by listener and accent within this task are nevertheless revealing.

The top panel of Fig. 2 presents only aggregate responses. In fact, listeners are not identical in their behavior. The bottom panels of Fig. 2 illustrate a three-way (tensor product) interaction between time, accent, and listener age. For ease of presentation, age is represented as a dichotomous variable, split into older (above the median age of thirty-seven) and younger (below the median), but all modeling was done with age as a continuous predictor. (Gender was not selected as a significant predictor in our models.)

The bottom panels show that younger listeners (left panel) consistently rate candidates more favorably than older listeners do (right panel), as observed in our parallel studies using retrospective ratings (Levon et al. 2021). They also show that the RP speaker is rated consistently more highly than the MLE speaker despite identical content. But the most important detail is that the response trajectories of younger and older listeners for MLE versus RP differ (for the age effect, $\chi^2 = 314.73$, $df = 20$, $p < 0.000$). The slopes of evaluation trajectories of the RP-speaking candidate show little difference across older and younger listeners (dashed lines in lower panels). This is not the case for evaluations of the MLE-speaking candidate (solid lines in lower panels). Younger listeners' response trajectory to MLE is similar to their reactions to RP, with the familiar linear monotonic increase over the course of the stimulus. Older listeners' evaluations of the MLE speaker remain almost completely flat until nearly 70 percent of the way through the stimulus. Only in the final quarter of the response do we see a slight increase in older listeners' average ratings. Possible interpretations of this absence of upward movement are discussed later. But this nearly flat distribution of real-time responses among older listeners indicates some form of existing biases and predispositions that are inhibiting dynamic updating based on unfolding evidence (cf. Campbell-Kibler 2009, 2011, Levon 2014).

Before turning to expert stimuli, we note that the general upward trajectory in Fig. 2 is observed in almost all of our results.[4] This may reflect a general reward for responding coherently in a job interview context, as all responses were produced with standard grammar, in a formal style, and with credible content for a law firm interview. (Appendix C shows that downward movement is not ruled out; Levon et al. (2020) also report a downward trajectory for a regional accent in a nonhiring context with the same experimental design.[5]) In our comparison to expert stimuli, we move beyond the upward direction of trajectories and compare trajectory shape.

Our analysis so far supports a model of impression formation that involves early categorization of speakers, based on prevailing British language ideologies. For younger listeners, this early categorization does not disrupt the process of updating, but for older listeners, categorizing a candidate as an MLE speaker appears to block the possibility that the individual is signaling worthiness for a job in a law firm, limiting the potential value of high-quality content.

4.2. Evaluations of expert stimuli. We included expert stimuli to examine whether signals of technical knowledge and legal expertise perturb listeners' prior expectations regarding the suitability of an RP or an MLE speaker for employment in the legal profession (see Berger et al. 1977). As before, we present model results of listeners' real-time evaluations of one of the two expert questions, for ease of presentation; Appendix C shows that results for the other expert question were comparable.

Figure 3 illustrates model-predicted results for evaluations of competence when speakers respond to the expertise-demanding question 'What is the difference between contract and tort?'. Models indicate that the only significant dynamic effect is a difference in how listeners evaluate the RP versus the MLE speaker over time. There is an additional fixed effect of speaker (showing an overall preference for the RP-speaking candidate), as well as significant tensor product interaction between listener age and MCPR. As before, gender was not selected as a significant predictor. For the purposes of our current discussion, we focus on the real-time differences in evaluations across speakers (see Appendix B for full model results).

As with the nonexpert questions, we find a basic upward trajectory across all evaluations here, toward higher evaluation as speaking continues. However, the evaluation trajectories of listener responses to expert questions differ from what we observed for nonexpert questions in two crucial ways.

First, the response trajectories for both RP and MLE in Fig. 3 are nonlinear. This differs significantly from the linear response trajectories we saw earlier for nonexpert questions.[6] This is confirmed by model results, which indicate an *EDF* value of 3.736 for the smooth associated with evaluations of the MLE speaker over time ($F = 40.306$,

[4] The trajectories plotted in Fig. 2 smooth out individual variation in the data. Inspection of individual responses reveals saccade-like patterns, with rightward movement on the response scale occurring as a series of successive jumps with plateaus in between. Statistical examination of trajectory curvature confirms that individual distributions are consistent with dynamic updating of evaluations rather than simply the implementation of a prior discrete categorization.

[5] Levon et al.'s (2020) result also suggests that the upward trajectory is not likely to be caused by the starting point of 50. Likewise, the long stimulus length (thirty seconds) makes it likely that the movement is not simply the time it takes to move the cursor toward 100. And finally, the dynamic nature of the responses may not be due simply to the instruction to attend to improvement or decline, as Levon et al. (2020) observed similar real-time dynamism with an instruction to simply judge 'professionalism'.

[6] To confirm a significant difference in trajectory shapes across expert and nonexpert questions, we built an additional set of GAMMs that examined evaluations of individual accents separately and compared response trajectories for expert versus nonexpert questions within each. These models replicate our overall findings, showing significant differences in response trajectories for expert versus nonexpert questions in MLE (for

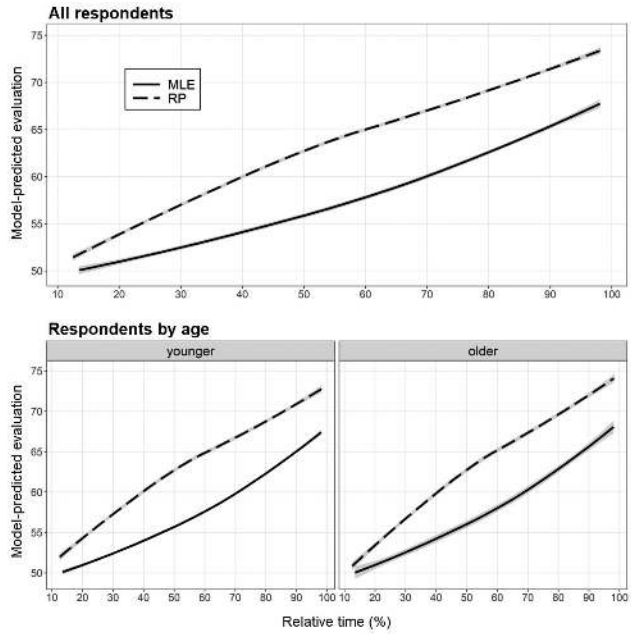

FIGURE 3. Model-predicted evaluations of responses to an expert question.

$p < 0.000$) and an *EDF* value of 4.031 ($F = 3.406$, $p = 0.004$) for the difference smooth associated with evaluations of the RP speaker over time.[7] (Recall that the further from 1 an *EDF* value is, the more nonlinear a response trajectory.) This indicates that while, as before, listeners update their impressions of speakers across the course of the stimulus, this updating is not at a constant rate. Instead, ratings for both the RP and the MLE speaker resemble polynomial trajectories, involving larger shifts in listener impressions at particular points in time.

Second, the shape of the polynomial evaluation trajectory for the RP speaker is the INVERSE of that found for the MLE speaker. In Fig. 3, the curve associated with RP is concave, characterized by a steep early rise in evaluative responses (approximately 15–55% of the stimulus) followed by a flatter period of upward incrementation (60–100% of the stimulus). The curve for MLE, in contrast, is convex, starting off with a long slow rise in evaluative responses (approximately 10–50% of the stimulus), followed by a much more rapid increase in the second half of the stimulus.[8]

We can see the difference in reactions to RP and MLE in the context of expert content more clearly in the two plots in Figure 4. On the left, we see that the difference in ratings over time for the two accents was linear for nonexpert stimuli (dashed line), but is dramatically nonlinear for expert stimuli (solid line). The right panel provides further information about the shape of the nonlinear response trajectories for expert content by

nonexpert: *EDF* = 1.000, $F = 49.36$, $p < 0.000$; for expert: *EDF* = 2.15, $F = 70.3$, $p = 0.000$) and in RP (for nonexpert: *EDF* = 1.000, $F = 159.8$, $p < 0.000$; for expert: *EDF* = 2.63, $F = 89.09$, $p < 0.000$).

[7] As before, treatment contrasts were used for model building, with MLE as the reference level for accent. The response trajectory for MLE therefore corresponds to the model smooth for the effect of time, while the response trajectory for RP corresponds to the difference smooth for time by accent. See n. 3 for details.

[8] These differences are not tied to feature density (see Appendix A): as with all questions, the divergence between RP and MLE ratings begins as soon as slider movement begins, approximately two seconds after the start of the stimulus.

showing the velocity of ratings of RP and MLE (Wojnowicz et al. 2009). The vertically higher dashed curve on the left of the graph indicates that ratings change faster for the RP speaker during the first half of the stimulus (concave curve). The higher solid line on the right of the graph indicates the opposite pattern for the MLE speaker, with more rating changes per second in the second half of the stimulus (convex curve).

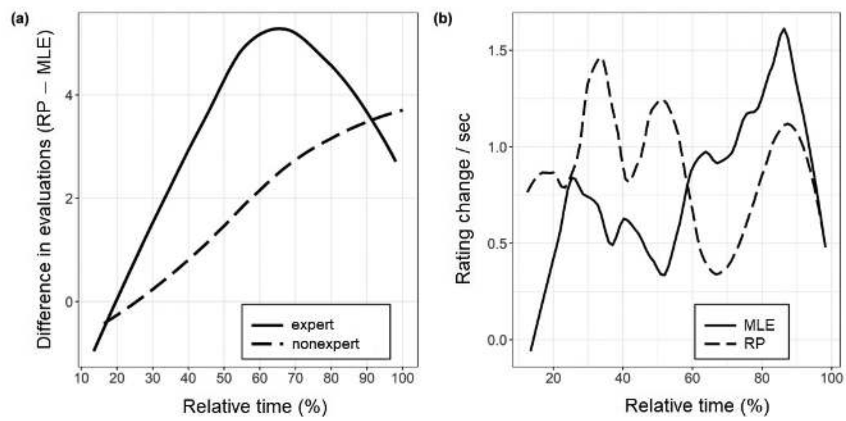

FIGURE 4. (a) Difference in mean model-predicted evaluations of the RP and MLE speakers for expert versus nonexpert responses. (b) Velocity of evaluative change over time for RP versus MLE for expert responses.

Before our discussion of these findings, we present one final set of data relevant to Figs. 3 and 4. The presence of nonlinear trajectories with different shapes for RP and MLE (concave and convex, respectively), despite identical lexical content, raises the possibility that listeners are judging perceived competence in relation to specific accent features (the type 3 prediction). To assess this, we developed a range of different dialect-density metrics, designed to test sensitivity to the presence of distinctive, nonstandard features in a given stimulus. A full discussion of dialect density is beyond the scope of the current article (Levon et al. 2020 discusses differences in retrospective ratings of individual speakers when accent density differs), but Figure 5 presents listeners' evaluative ratings alongside the relative density of accent features.

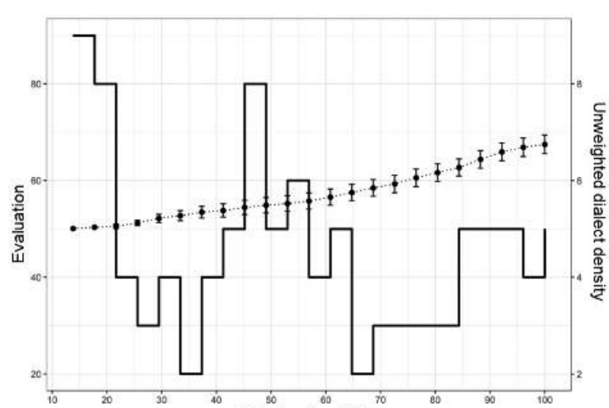

FIGURE 5. Observed mean ratings (and 95% confidence intervals) for the MLE speaker answering an expert question (dotted line). Solid line indicates the raw count of MLE accent features occurring within the previous three seconds of the stimulus.

The dotted line in Fig. 5 represents the mean observed evaluations (*y*-axis on the left) for the MLE speaker answering the expert question analyzed in Figs. 3 and 4. Superimposed over this curve is a step graph (solid line) indicating the number of MLE features (*y*-axis on the right) encountered at each time point in the stimulus. Scores are calculated in a rolling three-second window, such that a score at any given point of time indicates how many MLE features a listener encountered in the previous three seconds of the stimulus.

Figure 5 shows no clear correlation between changes in listener ratings and MLE dialect-density scores across the stimulus. While observed mean ratings increase steadily (if unevenly) across the course of the candidate's response, the presence of MLE features fluctuates, with peaks of nine and eight features at 15% and 45%, respectively, and valleys of three and two features at 25%, 35%, and 65%, respectively. We see no isomorphic changes in evaluations (e.g. a decrease in competence ratings in response to a rise in MLE dialect density). This suggests little evidence in the present data of listeners basing their real-time evaluative judgments directly on specific accent features (Austen & Campbell-Kibler 2022; though cf. Hesson & Shellgren 2015, Montgomery & Moore 2018). It is nevertheless worth noting that, in Fig. 5, the density of MLE features is higher in the first half of the utterance than it is in the second. It is therefore possible that the steeper rise in listeners' evaluations after the halfway mark is in some way linked to the generally more infrequent occurrence of MLE features in the latter portion of the response. Statistical modeling of the effect of MLE dialect density on listener evaluations did not confirm this. We ran multiple GAMM models on the MLE data only, which included different types of dialect-density measures. None of these showed a significant correlation between dialect density and listeners' evaluation trajectories. Thus, while the relative prevalence of dialect features may certainly play a role in listeners' overall evaluations (see Levon et al. 2021), the distribution of individual features within a stimulus does not appear to finely govern the observed response trajectories.

5. DISCUSSION. We discuss the implications of these results for real-time sociolinguistic perception in terms of three key findings: PRIOR BIAS in the initial response to the attitude object, universal DYNAMIC UPDATING of evaluations, and TRAJECTORY BIAS in the shape of dynamic updates.

5.1. PRIOR BIAS IN INITIAL REACTION. At the outset we outlined three possible types of real-time evaluative response based on the literature to date: absence of bias, where changes in evaluation do not track accent features but rather factors such as content (type 1); early broad categorization of speakers based on general recognition of the accent (type 2); and fine-grained fluctuations that track specific accent features or shibboleths (type 3).

All of our results show more support for type 2 evaluations than for the other types, with early broad differentiation of the RP and the MLE speaker within a few seconds of the audio stimuli (cf. Kraus et al. 2019). We argue that this early split in evaluations of the two speakers, often accompanied by a steeper positive slope of evaluation for RP, is the result of PRIOR BIAS—listeners' preexisting probabilistic expectations—in favor of RP. We concede that this claim is based on observed evaluations of one speaker per accent, but note the consistency of this pattern across questions and with prior research.

The absence of clear type 3 effects resembles Austen and Campbell-Kibler's (2022) finding that real-time responses do not consistently track linguistic cues. We also concur with their view that real-time monitoring likely involves very different cognitive processing from retrospective ratings. Unlike their findings, however, we do find evidence of global updating of evaluations, with clear effects of stimulus content and listener traits on overall trajectories; this may be due to our more goal-oriented task of assessing a speaker's likelihood of doing well in the interview. We share Austen and Campbell-Kibler's caution about the use of real-time tools to track sociolinguistic perception, but suggest that our results demonstrate some important social properties of listening.

As we will see in the sections that follow, early social categorization of a speaker plays a central role in evaluative processing. In both expert and nonexpert conditions, early categorization sets in motion a chain of perceptual effects—classifying and imbuing, to use Niedzielski and Preston's (2003) terminology—that condition the ensuing response trajectory. This finding is consistent with much prior work in sociolinguistic cognition and underscores the relevance of prior expectations for in-the-moment impression formation (e.g. Hay & Drager 2010, Walker & Hay 2011, Sumner et al. 2014).

5.2. Dynamic updating of priors. The early categorization and evaluation of RP and MLE did not remain stable for the duration of the listening event. All listeners progressively updated their evaluations, so prior bias did not fully determine the ensuing evaluative response profiles. Despite their early classification of speakers into social categories (and the concomitant imbuing of these classifications with attitudinal predispositions), listeners nevertheless integrated additional information about the speakers in the course of listening to them, and updated their evaluations accordingly. These unfolding details came from context (e.g. job interview, legal sector), content (e.g. use of technical vocabulary), and further factors (e.g. accent features in the speech stream).

In our data, this updating mostly results in increasingly positive ratings over time, for reasons discussed earlier. The only situation in which this updating seems nearly inactive is for older listeners when they are evaluating the MLE speaker in a nonexpert condition (Fig. 2 and Appendix C). There, we find a mostly flat response trajectory, where the older cohort's priors regarding likely characteristics of the MLE speaker remained largely unaffected by the utterance. In this case, we do not say that dynamic updating is absent, as we assume updating is a universal Bayesian property of real-time processing, but that updating is minimal here. There are a number of possible reasons for this: (i) older listeners start out with more established priors regarding MLE, lawyers, or both, so their evaluations are more resistant to dynamic updating; (ii) younger listeners have been exposed to a wider range of MLE exemplars, as MLE is a relatively new variety, so they can activate compatible alternatives more easily; or (iii) younger listeners may have fewer priors regarding the profession in question. Further studies would be needed to distinguish among these alternatives.

Our results provide a clear indication that sociolinguistic perception is about not only the activation of categories but also the ongoing and iterative process in which these activations are updated and dynamically reimbued with meaning over the course of social interaction (cf. the 'ideal adapter' framework: Kleinschmidt & Jaeger 2015, Kleinschmidt, Weatherholtz, & Jaeger 2018; socially weighted encoding: Sumner et al. 2014; and reentrant mapping: Johnson 2006).

5.3. Trajectory bias. The distinct behavior of older listeners illustrates our most striking finding regarding dynamic sociolinguistic processing. We use the term trajectory bias for this finding: that is, bias that is manifested in the trajectory of dynamic updating. The first example of this has been noted, namely, the minimal updating seen in the evaluation trajectory of older listeners. The other clear manifestation of trajectory bias was the overall difference in shape of response trajectories when listening to nonexpert and to expert stimuli.

In both expert and nonexpert conditions, we saw early categorization of the speakers based on their accent, but the former exhibited steady linear updating, while the latter exhibited inverse polynomial trajectory shapes for both the RP and the MLE speaker.[9] In the expert condition, participants exhibited a concave trajectory while listening to RP: evaluations rose early and quickly, almost automatically, with the slope of the rise tapering off in the latter half of the stimulus. For the MLE speaker, we found the opposite pattern: the listeners' priors regarding social attributes of an MLE speaker acted as a drag on judgments of his competence. The update of the MLE speaker accelerated only toward the end of the stimulus, with a final outcome that was significantly lower than that for the RP speaker ($t = 5.455$, $p < 0.000$). In simple Bayesian terms, it appears to take longer for the expert content of the answer to revise priors that anticipated a poor performance or poor knowledge. Even if the final judgments ultimately converge, the discrepancy in response trajectories places a distinctly greater burden of proof on the MLE speaker to demonstrate and sustain a 'good' response, and a greater need for sustained attention on the part of the listener.

The difference in evaluative reactions to expert and nonexpert question–response pairs may be due to distinct types and levels of contextual enrichment being activated, with expert questions priming listeners for additional, and potentially more stringent, status-linked expectations (Foschi 2000; see Levon et al. 2021 for full discussion).

In terms of real-time processing in the individual, how should we interpret this comprehensive (not feature-specific) delay when listeners encounter expert job interview responses in an MLE accent? As outlined in §2, we argue that the key to addressing this question lies in conceptualizing sociolinguistic perception as a process of DYNAMIC COMPETITION, that is, an iterative process of entertaining the probability of multiple possible outcomes before ultimately converging on a judgment. A framework of perception as dynamic competition developed in psychology (e.g. Freeman et al. 2008, Freeman et al. 2010; see also Squires 2014) has used a mouse-tracking paradigm to show that social category selection after exposure to a visual stimulus (e.g. a photo of a woman or a man) does not proceed in discrete stages, but rather via dynamic competition among multiple partially activated categories. Immediately following exposure to the stimulus, we partially activate multiple representations (e.g. woman and man) that are in continuous competition with one another until one of them ultimately wins out. With 'typical' stimuli, the cues to category membership can be so strong—and the competition process itself so rapid—that the selection process appears to be discrete and automatic. But atypical stimuli can uncover the process of dynamic competition (or, in Bayesian terms, change in probabilities).

Examining explicit attitudes with a similar mouse-tracking paradigm, Wojnowicz et al. (2009) showed asymmetric mouse trajectories when respondents were asked to select 'like' and 'dislike' for different racial categories. Although all respondents ultimately selected 'like' for both Black and White people, the mouse trajectories in the 'Black people' condition bear the hallmarks of greater dynamic competition, with greater overall curvature toward the 'dislike' side of the screen, greater spatial disorder ('wiggliness'), nonlinear velocity of mouse movements, and increased velocity toward

[9] The difference between linear and nonlinear response trajectories for nonexpert and expert questions, respectively, is reminiscent of discussions in the psychology literature on so-called ballistic versus nonballistic processing, where the former is unavoidable and automatic, while the latter is dynamic and subject to updating (see e.g. Brown & Heathcote 2005, 2008). We note this parallel here as a potentially interesting avenue for future research.

the end of the trajectories. Wojnowicz et al. (2009:1429) argue that, in this context, 'one attitude (whose supporting biases rise quickly in activation) [is] briefly prominent during early moments of forming an attitude choice, [while] a different attitude (whose supporting biases are stronger but rise in activation more slowly) takes hold during later moments of forming that same attitude choice'. Our real-time response trajectories in the expert question condition very closely resemble those identified by Freeman et al. (2008) and Wojnowicz et al. (2009), with the delayed resolution of rating for the MLE speaker indicating that both response alternatives ('doing well' and 'doing poorly') are simultaneously active and competing over the course of the stimulus.

Figure 6 schematically pulls together the three components of real-time sociolinguistic processing that we have discussed: indications of bias in the priors of listeners that emerge almost immediately during the listening event, a universal process of dynamic updating, and distinct types of trajectory bias corresponding to strength of priors and live competition among alternative ratings, or perhaps even stereotypical personae.

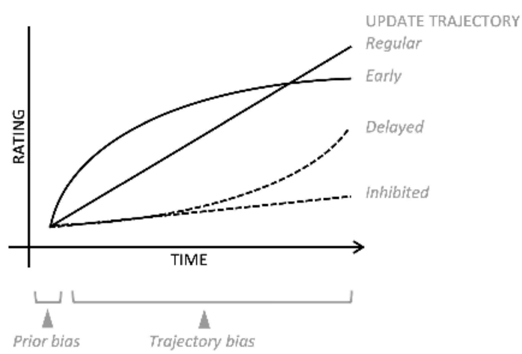

FIGURE 6. Schematic diagram of prior bias and different forms of trajectory bias in sociolinguistic processing. Dotted lines indicate trajectory types observed only for MLE.

Our findings reinforce the pivotal role of context in shaping the entire dynamic impression-formation process (Campbell-Kibler 2009, Hilton & Jeong 2019, Levon & Ye 2020). We operationalized context in two ways in this study. First, the situation of a job interview in a law firm provides an overriding frame for judgments of competence, presumably activating stereotypical expectations regarding the relationship between prestige, elite employment, and accents. Second, within this professional frame, the content manipulation between expert and nonexpert questions reveals further divergence depending on the contextual presence of technical expertise. In both cases, context has a significant impact, with notably different response trajectories to RP and MLE arising out of distinct status-linked expectations. These reported evaluations of RP and MLE are not proposed as representative of attitudes toward these varieties in all situations (if such global attitudes can even be said to exist), but rather as arising within specific contexts, and specifically within the understudied dimension of listening trajectory.

6. CONCLUSION. This study has looked at how we listen in real time. Our analyses of listeners' real-time evaluations of an MLE- and an RP-speaking candidate for a job in a UK law firm reveal several important patterns that have ramifications for how we model sociolinguistic perception, attitude formation, and the effects of linguistic bias.

Perception integrates social, cognitive, and contextual factors. In the present case, social factors are at play in the prior expectations that listeners bring to the event, cognitive factors take the form of universal dynamic updating of probabilistic expectations as new information is encountered, and contextual factors structure this new information (e.g. law, expert/nonexpert content). Together, these give shape to rating changes in real time, with prior bias observable within the first seconds of listening, followed by distinct forms of trajectory bias in how dynamic updating proceeds. In some cases, priors almost fully inhibit the potential for new information—for example, regarding technical expertise or professionalism—to revise stances. In others, they impose a considerable delay on updating. We suggest that the details of curvature and velocity of changes provide empirical support for a view of perception as dynamic, probabilistic competition rather than discrete activation, and a need to emphasize contextual situatedness in that body of research.

The findings confirm previous descriptions of evaluation as an active process of sense-making (Giles 2011), inference under uncertainty, and incremental adaptation (Johnson 2006, Sumner et al. 2014, Kleinschmidt & Jaeger 2015). Much of this research has been necessarily reliant on narrow phonetic or word-recognition paradigms, in order to understand the precise process of mapping acoustic cues to phonetic or lexical categories. As Kleinschmidt and Jaeger (2015:51) acknowledge: 'we have generally simplified our discussion by focusing on a few phonetic features. However, in everyday speech perception the generative model is highly complex, covering many categories, each cued by many acoustic features'. In the present work, we have extended these fine-grained models to holistic variability in the signal across naturalistic situations. The resulting diversity in update trajectories shows that dynamic updating plays out differently across social situations and content, drawing on listener priors about talker characteristics. Our empirical findings thus support the challenge of scaling up recent computational models in psycholinguistics, while also integrating those dynamic models into sociolinguistic understanding of language attitudes.

These findings also have direct implications for discrimination in real-world interactions. Even seemingly equitable outcomes (e.g. when RP and MLE attain a similar final rating) can hide inequitable trajectories of judgment along the way. Our finding of delayed updating for MLE implies that if an MLE speaker has the opportunity to speak for only a short time, or if the listener's attention is inconsistent, there may be insufficient time or content to shift the listener's prior expectations, a disadvantage that a standard speaker does not face. In other words, a greater burden of proof is placed on speakers of stigmatized accents: the 'jury is out' when they start speaking, increasing the probability that a minor slip or ambiguity may feed and promote a negative stereotype that is dynamically in play throughout.

APPENDIX A: QUESTION TEXTS WITH ACCENT ANNOTATIONS

VARIABLE	RP	MLE
PRICE vowel	[aɪ], [ʌɪ]	monophthongal, fronted [æ]
FACE vowel	[eɪ], [ɛɪ]	monophthongal [e]
GOAT vowel	[əʊ], [oʊ]	monophthongal [o]; backed onset
FOOT vowel	[ʊ], can be fronted	backed, raised [u]
coda /l/ ('l-vocalization')	velarized [ɫ]	[w]
glottal replacement of /t/	some in coda; not intervocalic	common in coda; some intervocalic; occasionally /k/ and /p/
/ð/ ('dh-stopping')	[ð]	[d̪], [d]
/ð/ ('TH-fronting')	[ð]	[v]
/θ/ ('th-fronting')	[θ]	[f]
onset /k/	[k]	backed, velarized [q]
onset /d/	[d]	[d̪]

(TABLE A1. *Continues*)

VARIABLE	RP	MLE
-in/*-ing*	more [ɪŋ]	more [ɪn], some [ɪŋk]
word-final *-er*/*-or*/*-a*	[əː]	[ʌː]
intrusive /r/	some use	less use
THOUGHT vowel	[ɔ]	raised [o] *(Nigerian)*
NURSE vowel	[ɜː]	fronted [ɛː] *(Nigerian)*
STRUT vowel	[ʌ]	[ɐ] *(Nigerian)*
onset /l/	clear [l]	velarized [ɫ] *(Nigerian)*

TABLE A1. Phonetic features observed for each accent in the recordings.

Note: Table A1 lists audibly distinctive features instantiated in the recordings, not just those documented in the literature, as less-described features can contribute to differentiating accents (Strycharczuk et al. 2020). Subtler phonetic cues were also present but not subjected to auditory coding. RP recordings were on average 1.76 seconds longer than MLE equivalents. The specific variety of MLE used by this London-born talker includes traces of Nigerian heritage. The boldface elements in the texts below highlight loci where a feature in the MLE recording contrasted audibly with the variants used in the RP recording, listed in Table A1.

NONEXPERT QUESTION–RESPONSE TEXTS

Why do you want to be a lawyer?

[MLE] I've wanted to be a l**aw**yer since **I** was a young ch**il**d. My fami**l**y used to j**ok**e that I was extreme**ly** g**oo**d at argu**ing** and I ought to be a l**aw**yer. S**o** **th**is des**i**re m**a**de me pursue a l**aw** **d**egree and **th**en **th**e LPC. It would be a gr**ea**t opportuni**t**y to work on a big **d**e**al** and e**x**perience the pr**o**cess. **I**'ve a**l**ways h**op**ed to w**or**k in a large law f**ir**m and be part of **th**e exc**i**tement and **th**e success here. **I** find **th**e career very appea**l**ing and I would l**o**ve to w**or**k for **th**e f**ir**m.

What attracts you about working in law?

[MLE] Law is interest**ing** and I enjoyed m**y** st**u**dies. I **c**an**'t** pretend **th**at money isn'**t** a fact**or**. We a**ll** come out of universi**ty** with significant debt, and a career in law allows a g**oo**d **q**uali**ty** of l**i**fe and financia**l** securi**ty**. I'm also very interested in **th**e internationa**l** aspects of w**or**king in law. I've a**l**ways dreamed of w**or**k**ing** in Amer-i**ca**, and **th**is could be **th**e first step towards internationa**l** career experience too. **I** w**or**k hard and I fee**l** m**y** st**y**le of w**or**king would be sui**t**ed to law, and **th**e t**y**pe of internationa**l** w**or**k tha**t** is **do**ne.

EXPERT QUESTION–RESPONSE TEXTS

Explain the difference between contract and tort.

[MLE] **C**ontrac**t** l**aw** invo**l**ves l**aw**s **th**at **d**eal wi**th** **c**ontrac**t**s between peop**le**, and tor**t** l**aw** is where someone **d**oes s**o**meth**ing** wrong. You can of **c**ourse do s**o**me**thing** wrong in contrac**t** law as we**ll**, bu**t** **that** invo**l**ves breach of **c**ontract l**aw** ra**th**er **th**an tor**t** l**aw**. In our tr**aining**, we **c**overed **c**ontract law first, and **th**en tor**t** law and **c**ompared and **c**ontrasted c**a**ses. **Th**ere are similari**t**ies across **the** c**a**ses, bu**t** a**l**so **d**istinct se**t**s of princip**le**s in the different situ**a**tions as **o**n**l**y **o**ne invo**l**ves **c**ontract agreements.

How has the Human Rights Act 1998 affected law in this country?

[MLE] **Th**e Human R**ig**h**t**s Act is an impor**t**an**t** piece of legislation. **Th**e requ**i**remen**t** for a pr**i**va**t**e and family l**i**fe means **th**at s**o**me crimina**l**s **c**an'**t** be depor**t**ed for a long t**i**me and **that** prisoners have **the** r**igh**t to vo**t**e. Some o**f** **th**ese requ**i**remen**t**s are problematic. You **c**ould argue **that** crimina**l**s and prisoners shouldn'**t** have **th**ese r**igh**ts. So i**t** has d**o**ne s**o**me g**oo**d bu**t** I do disagree with some aspec**t**s of i**t**. Bu**t** i**t** has been usefu**l** to have human r**igh**ts se**t** ou**t** in one pl**a**ce **th**is w**a**y.

APPENDIX B: FULL MODEL RESULTS

PARAMETRIC TERM	EST	*t*	*p*-VALUE
(intercept)	55.047	186.696	< 0.000
Accent (RP)	3.843	8.885	< 0.000
SMOOTH TERM	*EDF*	*F*	*p*-VALUE
s (Age)	2.063	66.362	< 0.000
s (MCPR)	2.428	20.610	< 0.000
s (Time)	1.000	26.277	< 0.000
s (Age) : Accent (RP)	2.953	27.035	< 0.000
s (MCPR) : Accent (RP)	1.001	66.786	< 0.000
s (Time) : Accent (RP)	1.000	4.202	0.041
ti (Age, MCPR)	7.861	7.766	< 0.000
ti (Time, MCPR)	1.000	2.407	0.121

(TABLE A2. *Continues*)

SMOOTH TERM	*EDF*	*F*	*p*-VALUE
ti (Time, Age)	1.903	5.374	0.003
ti (Age, MCPR) : Accent (RP)	8.861	46.133	< 0.000
ti (Time, Age) : Accent (RP)	1.000	5.176	0.023
ti (Time, MCPR) : Accent (RP)	1.879	1.200	0.293
ti (Time, Age, MCPR)	1.708	0.461	0.601
ti (Time, Respondent)	186.848	1.081	< 0.000

Adjusted R^2 = 0.384, deviance explained = 43%, *N* = 2,984, respondents = 153

TABLE A2. GAMM for nonexpert question.

PARAMETRIC TERM	EST	*t*	*p*-VALUE
(intercept)	57.662	78.789	< 0.000
Accent (RP)	5.590	5.455	< 0.000
SMOOTH TERM	*EDF*	*F*	*p*-VALUE
s (Age)	1.000	0.042	0.839
s (MCPR)	1.000	6.947	0.008
s (Time)	3.736	40.406	< 0.000
s (Age) : Accent (RP)	1.000	0.000	0.992
s (MCPR) : Accent (RP)	2.102	2.028	0.108
s (Time) : Accent (RP)	4.031	3.406	0.004
ti (Age, MCPR)	2.543	2.908	0.030
ti (Time, MCPR)	2.813	1.727	0.117
ti (Time, Age)	1.003	0.094	0.763
ti (Age, MCPR) : Accent (RP)	1.000	7.203	0.007
ti (Time, Age) : Accent (RP)	1.000	0.216	0.642
ti (Time, MCPR) : Accent (RP)	1.496	0.745	0.557
ti (Time, Age, MCPR)	4.599	0.810	0.580

Adjusted R^2 = 0.278, deviance explained = 27.4%, *N* = 3,721, respondents = 157
Autocorrelation error (rho) = 0.919

TABLE A3. GAMM for expert question.

APPENDIX C: MODEL PREDICTIONS

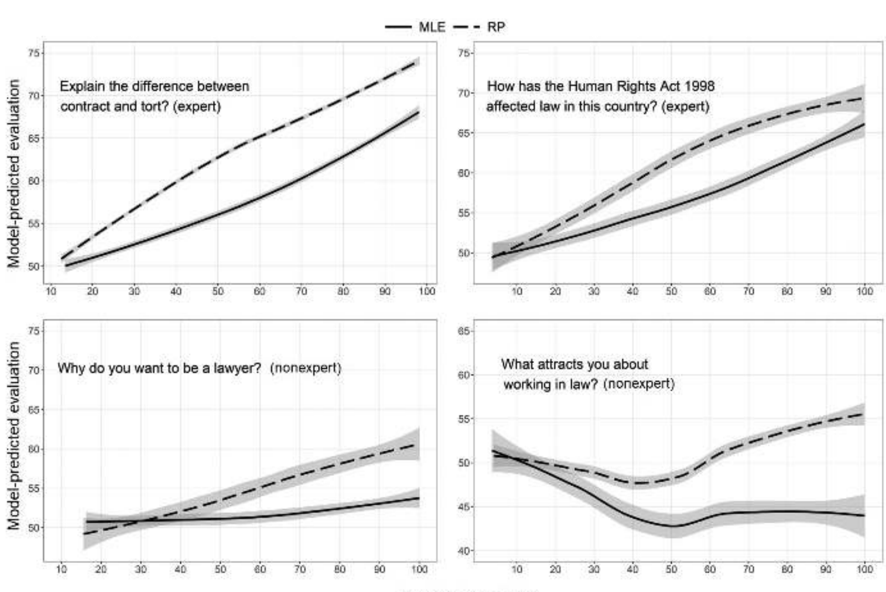

FIGURE A1. Graphical comparison of model predictions for evaluations by older respondents to two expert questions (top row) and two nonexpert questions (bottom row).

REFERENCES

Afifi, Walid A., and Judith L. Weiner. 2004. Toward a theory of motivated information management. *Communication Theory* 14.167–90. DOI: 10.1111/j.1468-2885.2004.tb00310.x.

Austen, Martha, and Kathryn Campbell-Kibler. 2022. Real-time speaker evaluation: How useful is it and what does it measure? *Language* 98.e108–e130. DOI: 10.1353/lan.2022.0000.

Baayen, R. Harald; Jacolien van Rij; Cecile de Cat; and Simon Wood. 2018. Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. *Mixed-effects regression models in linguistics*, ed. by Dirk Speelman, Kris Heylen, and Dirk Geeraerts, 49–69. Cham: Springer. DOI: 10.1007/978-3-319-69830-4_4.

Babel, Molly. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40.177–89. DOI: 10.1016/j.wocn.2011.09.001.

Babel, Molly, and Jamie Russell. 2015. Expectations and speech intelligibility. *The Journal of the Acoustical Society of America* 137.2823–33. DOI: 10.1121/1.4919317.

Berger, Joseph; M. Hamit Fisek; Robert Norman; and Morris Zelditch, Jr. 1977. *Status characteristics and social interaction: An expectation states approach*. New York: Elsevier.

Blanden, Jo; Paul Gregg; and Stephen Machin. 2005. Educational inequality and intergenerational mobility. *What's the good of education? The economics of education in the UK*, ed. by Stephen Machin and Anna Vignoles, 99–114. Princeton, NJ: Princeton University Press.

Brown, Scott D., and Andrew Heathcote. 2005. A ballistic model of choice response time. *Psychological Review* 112.117–28. DOI: 10.1037/0033-295X.112.1.117.

Brown, Scott D., and Andrew Heathcote. 2008. The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology* 57.153–78. DOI: 10.1016/j.cogpsych.2007.12.002.

Burnett, Heather. 2019. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy* 42.419–50. DOI: 10.1007/s10988-018-9254-y.

Cabinet Office. 2011. Opening doors, breaking barriers: A strategy for social mobility. Online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/61964/opening-doors-breaking-barriers.pdf.

Campbell-Kibler, Kathryn. 2008. I'll be the judge of that: Diversity in social perceptions of (ING). *Language in Society* 37.637–59. DOI: 10.1017/S0047404508080974.

Campbell-Kibler, Kathryn. 2009. The nature of sociolinguistic perception. *Language Variation and Change* 21.135–56. DOI: 10.1017/S0954394509000052.

Campbell-Kibler, Kathryn. 2011. Intersecting variables and perceived sexual orientation in men. *American Speech* 86.52–68. DOI: 10.1215/00031283-1277510.

Cardoso, Amanda; Erez Levon; Devyani Sharma; Dominic Watt; and Yang Ye. 2019. Inter-speaker variation and the evaluation of British English accents in employment contexts. *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)*, Melbourne, 1615–19. Online: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_1664.pdf.

Cargile, Aaron C., and Howard Giles. 1997. Understanding language attitudes: Exploring listener affect and identity. *Language and Communication* 17.195–217. DOI: 10.1016/S0271-5309(97)00016-5.

Cargile, Aaron C.; Howard Giles; Ellen B. Ryan; and James J. Bradac. 1994. Language attitudes as a social process: A conceptual model and new directions. *Language and Communication* 14.211–36. DOI: 10.1016/0271-5309(94)90001-9.

Casasanto, Daniel, and Gary Lupyan. 2015. All concepts are ad hoc concepts. *The conceptual mind: New directions in the study of concepts*, ed. by Eric Margolis and Stephen Laurence, 543–66. Cambridge, MA: MIT Press.

Chaiken, Shelly, and Durairaj Maheswaran. 1994. Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology* 66.460–73. DOI: 10.1037/0022-3514.66.3.460.

Cheshire, Jenny; Sue Fox; Paul Kerswill; and Eivind Torgersen. 2008. Ethnicity, friendship network, and social practices as the motor of dialect change: Linguistic innovation in London. *Sociolinguistica* 22.1–23. DOI: 10.1515/9783484605299.1.

Cheshire, Jenny; Paul Kerswill; Sue Fox; and Eivind Torgersen. 2011. Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15.151–96. DOI: 10.1111/j.1467-9841.2011.00478.x.

Clark, Gregory. 2014. *The son also rises: Surnames and the history of social mobility*. Princeton, NJ: Princeton University Press.

Cooper, Robert L., and Joshua A. Fishman. 1974. The study of language attitudes. *International Journal of the Sociology of Language* 1974(3).5–20. DOI: 10.1515/ijsl.1974.3.5.

Coupland, Nikolas, and Hywel Bishop. 2007. Ideologised values for British accents. *Journal of Sociolinguistics* 11.74–93. DOI: 10.1111/j.1467-9841.2007.00311.x.

D'Onofrio, Annette. 2015. Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics* 19.241–56. DOI: 10.1111/josl.12115.

Drager, Katie, and M. Joelle Kirtley. 2016. Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. *Awareness and control in sociolinguistic research*, ed. by Anna Babel, 1–24. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139680448.003.

Dragojevic, Marko, and Howard Giles. 2014. The reference frame effect: An intergroup perspective on language attitudes. *Human Communication Research* 40.91–111. DOI: 10.1111/hcre.12017.

Dunton, Bridget C., and Russell H. Fazio. 1997. An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin* 23.316–26. DOI: 10.1177/0146167297233009.

Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41.87–100. DOI: 10.1146/annurev-anthro-092611-145828.

Fabricius, Anne. 2005. Mobility, contact and an accent norm: The case of Received Pronunciation. *The consequences of mobility*, ed. by Bent Preisler, Anne Fabricius, Hartmut Haberland, Susanne Kjaerbeck, and Karen Risager, 120–34. Roskilde: Roskilde University. Online: https://ojs.ruc.dk/index.php/mobility/article/view/5342.

Foschi, Martha. 2000. Double standards for competence: Theory and research. *Annual Review of Sociology* 26.21–42. DOI: 10.1146/annurev.soc.26.1.21.

Fox, Sue. 2015. *The New Cockney: New ethnicities and adolescent speech in the traditional East End of London*. Basingstoke: Palgrave.

Frank, Michael C., and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336.998. DOI: 10.1126/science.1218633.

Freeman, Jonathan B.; Nalini Ambady; Nicholas O. Rule; and Kerri L. Johnson. 2008. Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General* 137.673–90. DOI: 10.1037/a0013875.

Freeman, Jonathan B.; Kristin Pauker; Evan P. Apfelbaum; and Nalini Ambady. 2010. Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology* 46.179–85. DOI: 10.1016/j.jesp.2009.10.002.

Friedman, Sam; Daniel Laurison; and Andrew Miles. 2015. Breaking the 'class' ceiling? Social mobility into Britain's elite occupations. *The Sociological Review* 63.259–89. DOI: 10.1111/1467-954X.12283.

Gal, Susan. 2016. Sociolinguistic differentiation. *Sociolinguistics: Theoretical debates*, ed. by Nikolas Coupland, 113–35. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781107449787.006.

Getz, Laura M., and Joseph C. Toscano. 2021. The time-course of speech perception revealed by temporally-sensitive neural measures. *WIREs Cognitive Science* 12:e1541. DOI: 10.1002/wcs.1541.

Giles, Howard. 1970. Evaluative reactions to accents. *Educational Review* 22.211–27. DOI: 10.1080/0013191700220301.

Giles, Howard. 2011. Language attitudes: The role of motivated information management. *Canadian Issues*, Autumn, 28–34.

Giles, Howard, and Mikaela L. Marlow. 2011. Theorizing language attitudes: Existing frameworks, an integrative model, and new directions. *Annals of the International Communication Association* 35.161–97. DOI: 10.1080/23808985.2011.11679116.

Giles, Howard, and Tamara Rakić. 2014. Language attitudes: Social determinants and consequences of language variation. *The Oxford handbook of language and social psychology*, ed. by Thomas Holtgraves, 11–26. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199838639.013.030.

Harrington, Jonathan; Felicitas Kleber; Ulrich Reubold; Florian Schiel; and Mary Stevens. 2018. Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in Cognitive Science* 10.707–28. DOI: 10.1111/tops.12329.

Hay, Jennifer. 2018. Sociophonetics: The role of words, the role of context, and the role of words in context. *Topics in Cognitive Science* 10.696–706. DOI: 10.1111/tops.12326.

Hay, Jennifer, and Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48.865–92. DOI: 10.1515/ling.2010.027.

Hay, Jennifer; Katie Drager; and Andy Gibson. 2018. Hearing *r*-sandhi: The role of past experience. *Language* 94.360–404. DOI: 10.1353/lan.2018.0020.

Hesson, Ashley, and Madeline Shellgren. 2015. Discourse marker *like* in real time: Characterizing the time-course of sociolinguistic impression formation. *American Speech* 90.154–86. DOI: 10.1215/00031283-3130313.

Hilton, Katherine, and Sunwoo Jeong. 2019. The role of context in sociolinguistic perception. *Linguistics Vanguard* 5:20180069. DOI: 10.1515/lingvan-2018-0069.

Jacewicz, Ewa; Robert A. Fox; and Joseph Salmons. 2007. Vowel duration in three American English dialects. *American Speech* 82.367–85. DOI: 10.1215/00031283-2007-024.

Johnson, Keith. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34.485–99. DOI: 10.1016/j.wocn.2005.08.004.

Johnson, Keith; Elizabeth A. Strand; and Mariapaola D'Imperio. 1999. Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics* 27.359–84. DOI: 10.1006/jpho.1999.0100.

Kerswill, Paul, and Ann Williams. 2002. 'Salience' as an explanatory factor in language change: Evidence from dialect levelling in urban England. *Language change: The interplay of internal, external and extra-linguistic factors*, ed. by Mari C. Jones and Edith Esch, 81–110. Berlin: De Gruyter. DOI: 10.1515/9783110892598.81.

Kim, Jonny, and Katie Drager. 2018. Rapid influence of word-talker associations on lexical access. *Topics in Cognitive Science* 10.775–86. DOI: 10.1111/tops.12351.

Kircher, Ruth, and Sue Fox. 2019. Multicultural London English and its speakers: A corpus-informed discourse study of standard language ideology and social stereotypes. *Journal of Multilingual and Multicultural Development* 42.792–810. DOI: 10.1080/01434632.2019.1666856.

Kleinschmidt, Dave F., and T. Florian Jaeger. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review* 122.148–203. DOI: 10.1037/a0038695.

Kleinschmidt, Dave F.; Kodi Weatherholtz; and T. Florian Jaeger. 2018. Sociolinguistic perception as inference under uncertainty. *Topics in Cognitive Science* 10.818–34. DOI: 10.1111/tops.12331.

Kraus, Michael W.; Brittany Torrez; Jun Won Park; and Fariba Ghayebi. 2019. Evidence for the reproduction of social class in brief speech. *Proceedings of the National Academy of Sciences* 116.22998–23003. DOI: 10.1073/pnas.1900500116.

Kristiansen, Tore. 2009. The macro-level social meanings of late-modern Danish accents. *Acta Linguistica Hafniensia* 41.167–92. DOI: 10.1080/03740460903364219.

Kristiansen, Tore. 2010. Attitudes, ideology and awareness. *The SAGE handbook of sociolinguistics*, ed. by Ruth Wodak, Barbara Johnstone, and Paul Kerswill, 265–78. London: SAGE.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William; Sharon Ash; Maya Ravindranath; Tracey Weldon; Maciej Baranowski; and Naomi Nagy. 2011. Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* 15.431–63. DOI: 10.1111/j.1467-9841.2011.00504.x.

Lev-Ari, Shiri; Emily Ho; and Boaz Keysar. 2018. The unforeseen consequences of interacting with non-native speakers. *Topics in Cognitive Science* 10.835–49. DOI: 10.1111/tops.12325.

Levon, Erez. 2014. Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society* 43.539–66. DOI: 10.1017/S0047404514000554.

Levon, Erez, and Isabelle Buchstaller. 2015. Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing. *Language Variation and Change* 27.319–48. DOI: 10.1017/S0954394515000149.

Levon, Erez; Isabelle Buchstaller; and Adam Mearns. 2020. Towards an integrated model of perception: Linguistic architecture and the dynamics of sociolinguistic cognition. *Advancing socio-grammatical variation and change: In honour of Jenny Cheshire*, ed. by Karen Beaman, Isabelle Buchstaller, Susan Fox, and James Walker, 32–54. London: Routledge.

Levon, Erez; Devyani Sharma; Amanda Cardoso; Yang Ye; and Dominic Watt. 2020. Attitudes to accents in Britain: Ideologies, phonetic detail and the reproduction of accent bias. Paper presented at the annual meeting of the Linguistic Society of America, New Orleans, January 2020.

Levon, Erez; Devyani Sharma; and Dominic Watt. 2017–2021. Accent bias and fair access in Britain. UK Economic and Social Research Council (ESRC). Online: https://accentbiasbritain.org/.

Levon, Erez; Devyani Sharma; Dominic Watt; Amanda Cardoso; and Yang Ye. 2021. Accent bias and perceptions of professional competence in England. *Journal of English Linguistics* 49.355–88. DOI: 10.1177/00754242211046316.

Levon, Erez, and Yang Ye. 2020. Language, indexicality and gender ideologies: Contextual effects on the perceived credibility of women. *Gender and Language* 14.123–51. DOI: 10.1558/genl.39235.

Macrae, C. Neil, and Galen V. Bodenhausen. 2001. Social cognition: Categorical person perception. *British Journal of Psychology* 92.239–55. DOI: 10.1348/000712601162059.

Mason, Rowena. 2013. David Cameron admits ministers must 'do far more' to increase social mobility. *The Guardian*, 14 November 2013. Online: https://www.theguardian.com/society/2013/nov/14/david-cameron-social-mobility-major.

McKenzie, Robert M., and Erin Carrie. 2018. Implicit-explicit attitudinal discrepancy and the investigation of language attitude change in progress. *Journal of Multilingual and Multicultural Development* 39.830–44. DOI: 10.1080/01434632.2018.1445744.

Montgomery, Chris, and Emma Moore. 2018. Evaluating S(c)illy voices: The effects of salience, stereotypes, and co-present language variables on real-time reactions to regional speech. *Language* 94.629–61. DOI: 10.1353/lan.2018.0038.

Mugglestone, Lynda. 2007. *Talking proper: The rise of accent as social symbol*. Oxford: Oxford University Press.

Niedzielski, Nancy A. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18.62–85. DOI: 10.1177/0261927X99018001005.

Niedzielski, Nancy A., and Dennis R. Preston. 2003. *Folk linguistics*. Berlin: De Gruyter. DOI: 10.1515/9783110803389.

Pantos, Andrew. 2019. Implicitness, automaticity, and consciousness in language attitudes research: Are they related and how do we characterize them? *Linguistics Vanguard* 5:20180007. DOI: 10.1515/lingvan-2018-0007.

Pharao, Nicolai; Marie Maegaard; Janus Spindler Møller; and Tore Kristiansen. 2014. Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society* 43.1–31. DOI: 10.1017/S0047404513000857.

Preston, Dennis R. 1996. Whaddayaknow? The modes of folk linguistic awareness. *Language Awareness* 5.40–74. DOI: 10.1080/09658416.1996.9959890.

Preston, Dennis R. 2010. Variation in language regard. *Empirische Evidenzen und theoretische Passungen sprachlicher Variation*, ed. by Peter Gilles, Joachim Scharloth, and Evelyn Zeigler, 7–27. Frankfurt: Peter Lang.

Preston, Dennis R. 2019. How to trick respondents into revealing implicit attitudes—talk to them. *Linguistics Vanguard* 5:20180006. DOI: 10.1515/lingvan-2018-0006.

R CORE TEAM. 2020. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: https://www.r-project.org/.

SHARMA, DEVYANI; EREZ LEVON; and YANG YE. 2021. 50 years of British accent bias: Stability and lifespan change in attitudes to accents. *English World-Wide* 43.135–66. DOI: 10.1075/eww.20010.sha.

SHAW, GEORGE BERNARD. 1916. *Pygmalion*. New York: Brentano.

SHERIDAN, THOMAS. 1762. *A course of lectures on elocution: Together with two dissertations on language; and some other tracts relative to those subjects*. London: W. Strahan.

SILVERSTEIN, MICHAEL. 1981. The limits of awareness. (Sociolinguistic working paper 84.) Austin, TX: Southwest Educational Development Laboratory. Online: https://eric.ed.gov/?id=ED250941.

SOCIAL MOBILITY COMMISSION. 2019. *State of the nation 2018–19: Social mobility in Great Britain*. Online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/798404/SMC_State_of_the_Nation_Report_2018-19.pdf.

SÓSKUTHY, MÁRTON. 2017. Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. (White Rose research online working paper.) York: University of York. Online: http://eprints.whiterose.ac.uk/113858/.

SQUIRES, LAUREN. 2014. Processing, evaluation, knowledge: Testing the perception of English subject–verb agreement variation. *Journal of English Linguistics* 42.144–72. DOI: 10.1177/0075424214526057.

STRYCHARCZUK, PATRYCJA; MANUEL LÓPEZ-IBÁÑEZ; GEORGINA BROWN; and ADRIAN LEEMANN. 2020. General Northern English: Exploring regional variation in the North of England with machine learning. *Frontiers in Artificial Intelligence* 3:48. DOI: 10.3389/frai.2020.00048.

SUMNER, MEGHAN; SEUNG KYUNG KIM; ED KING; and KEVIN B. MCGOWAN. 2014. The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology* 4:1015. DOI: 10.3389/fpsyg.2013.01015.

VAN RIJ, JACOLIEN; MARTIJN WIELING; R. HARALD BAAYEN; and HEDDERIK VAN RIJN. 2020. itsadug: Interpreting time series and autocorrelated data using GAMMs. R package. Online: https://cran.r-project.org/web/packages/itsadug/index.html.

WAGNER, SUZANNE EVANS, and ASHLEY HESSON. 2014. Individual sensitivity to the frequency of socially meaningful linguistic cues affects language attitudes. *Journal of Language and Social Psychology* 33.651–66. DOI: 10.1177/0261927X14528713.

WALKER, ABBY, and JEN HAY. 2011. Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology* 2.219–37. DOI: 10.1515/labphon.2011.007.

WATSON, KEVIN, and LYNN CLARK. 2013. How salient is the NURSE~SQUARE merger? *English Language & Linguistics* 17.297–323. DOI: 10.1017/S136067431300004X.

WATSON, KEVIN, and LYNN CLARK. 2015. Exploring listeners' real-time reactions to regional accents. *Language Awareness* 24.38–59. DOI: 10.1080/09658416.2014.882346.

WOJNOWICZ, MICHAEL T.; MELISSA J. FERGUSON; RICK DALE; and MICHAEL J. SPIVEY. 2009. The self-organization of explicit attitudes. *Psychological Science* 20.1428–35. DOI: 10.1111/j.1467-9280.2009.02448.x.

WOOD, SIMON N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society B: Statistical Methodology* 73.3–36. DOI: 10.1111/j.1467-9868.2010.00749.x.

WOOD, SIMON N. 2017. *Generalized additive models: An introduction with R*. 2nd edn. Boca Raton: Chapman and Hall/CRC.

YAEGER-DROR, MALCAH. 1993. Linguistic analysis of dialect 'correction' and its interaction with cognitive salience. *Language Variation and Change* 5.189–224. DOI: 10.1017/S0954394500001460.

YE, YANG; EREZ LEVON; DEVYANI SHARMA; DOMINIC WATT; and AMANDA CARDOSO. 2020. Mind the accent: Effects of bias reduction interventions on the evaluation of job candidates with accents in the U.K. Poster presented at the annual conference of the Society for Personality and Social Psychology, New Orleans, February 2020.

Levon
Center for the Study of Language and Society
Muesmattstrasse 45
3012 Bern, Switzerland
[erez.levon@unibe.ch] (Levon)
[d.sharma@qmul.ac.uk] (Sharma)
[y.ye@greenwich.ac.uk] (Ye)

[Received 11 February 2021;
revision invited 10 October 2021;
revision received 28 February 2022;
accepted pending revisions 27 April 2022;
revision received 6 June 2022;
accepted 9 June 2022]