INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

The different roles of expectations in phonetic and lexical processing

Shiri Lev-Ari[1], *Robin Dodsworth*[2], *Jeff Mielke*[2], *Sharon Peperkamp*[3]

[1]Royal Holloway, University of London, UK
[2]North Carolina State University, USA
[3]Laboratoire de Sciences Cognitives et Psycholinguistique (ENS – PSL University, EHESS, CNRS), France
shiri.lev-ari@rhul.ac.uk, robin_dodsworth@ncsu.edu, jimielke@ncsu.edu, Sharon.peperkamp@ens.fr

Abstract

The way people speak reflects their demographic background. Listeners exploit this contingent variation and make use of information about speakers' background to process their speech. Evidence for this comes from both phonetic and lexical tasks, and the two are assumed to tap into the same mechanism and provide equivalent results. Curiously, this assumption has never been tested. Additionally, while it has been established that expectations can influence language processing in general, the role of individual differences in susceptibility to this influence is relatively unexplored. We investigate these two questions in the context of Southern and General American speech varieties in the USA. We show that phonetic and lexical tasks are not equivalent, and furthermore, that the two are driven by mechanisms that are sensitive to different individual variables: while performance at the lexical level is influenced by implicit bias, performance at the phonetic level is influenced by working memory. These results thus change our understanding of how expectations influence processing, and have implications for how to conduct and interpret studies on the topic.

Index Terms: speech perception, expectations, implicit bias, working memory, autism spectrum quotient, lexical representations, phonetic representations

1. Introduction

1.1. The influence of expectations on speech perception

Speakers of different age, ethnicity, socioeconomic status, or geographical region often pronounce words differently [e.g., 1-3]. Speakers also use language to project their identity and express affiliation with certain social groups compared with others. Consequently, speakers with different political opinions or speakers who differ in the degree to which they affiliate with a certain group might also differ in the way they produce certain phonemes [e.g., 4-5].

Ideal listeners would make use of this information during language processing, as accounting for these variations can reduce ambiguity and facilitate comprehension. Indeed, previous research suggests that listeners rely on this information. Thus, it has been shown that listeners' speech perception is influenced by speakers' age, geographical region, ethnicity, and socioeconomic background [6-15]. While such reliance on expectations is likely to often be beneficial, expectations are sometimes inaccurate or informed by biases. In these cases, they could distort speech perception.

So far, the influence of expectations on speech perception has been mostly investigated with one of two types of paradigms: vowel matching, and lexical interpretation. In the typical vowel matching paradigm, participants hear a speaker produce a word containing a target vowel, and then indicate which vowel on a provided vowel continuum best matches the token they heard [8, 13]. In lexical paradigms, participants are presented with a potentially ambiguous word that could fit either word of a minimal pair, depending on the speaker's accent. Participants' lexical interpretation is tested to determine their decoding of the target vowel [6-7, 9-11, 15]. Both types of paradigms have yielded results that demonstrate an effect of expectations, such that performance differed when expectations regarding speakers' identities differed. Importantly, any difference in interpretation, in any direction, was interpreted as support for the role of expectations. A theory regarding the influence of expectation on speech processing, however, should make more precise predictions.

Most researchers in the field assume that representation includes exemplars that are stored together with social information [16-17]. Knowledge of certain social properties of the speaker lead to greater activation of tokens from speakers that share these social properties. Consequently, perception would be drawn towards these stored tokens, leading to exaggeration. That is, production would be perceived as more stereotypical than it is. For example, listeners are likely to perceive /ɪ/ as more raised and fronted than it is if the speaker is presented as Australian rather than New Zealander, because exemplars from Australian speakers, which tend to be more raised and fronted, would be activated more during processing. Indeed, studies using the vowel matching paradigm often report such an effect of exaggeration [8, 13]. In contrast, in a lexical task, expectations that a speaker would produce a word differently from a reference pronunciation such that it might sound like another word in the reference accent should lead listeners to compensate for the deviation from the reference accent. For example, expecting a speaker to omit [t] at the end of the word *belt* should push the listener upon hearing *bell* towards interpreting the input as if it had a [t], namely, *belt*. We'll term this effect compensation. In line with this hypothesis, most studies using a lexical paradigm have found an effect of compensation [7, 11-12, 15, but cf. 10].

The current study is the first to test participants using both paradigms and thus examine whether the two paradigms measure the same underlying effect. We hypothesize that the same expectations lead to an exaggeration effect in the vowel matching task but a compensation effect in the lexical task.

http://dx.doi.org/10.21437/Interspeech.2019-1795

1.2. Individual differences in language processing

It is uncontroversial that individual differences in cognitive abilities or attitudes can influence language processing. Nonetheless, previous literature on the effect of expectations on speech perception has only examined individual differences that are due to differences in expectations. For example, the lesser influence of tagging a speaker as Australian vs New Zealander for New Zealanders of lower socio-economic status was interpreted as driven by lesser familiarity with the stereotypical features of Australian accent [8]. Here we test whether individual differences in cognitive skills and attitude can influence individuals' susceptibility to the effect of expectations on speech perception by influencing how language is processed and information is integrated. This investigation will also enable us to examine whether the two paradigms, that tap phonetic and lexical processing, respectively, are similarly influenced by these individual differences.

1.2.1. Working Memory

The influence of expectations on speech perception requires integration of social information with the speech input. Various studies have shown that such integration can be effortful, and relies on Working Memory (WM). For example, WM has been shown to influence readers' ability to integrate animacy and syntactic information [18-19]. Furthermore, WM has been shown to modulate the influence of speaker-expectations on processing at the semantic level, such that individuals with higher WM interpret speech differently depending on speaker's identity, whereas the interpretation of listeners with lower WM tends to be invariant across contexts [20]. We test whether also at the phonological level, individuals with higher WM are more influenced by expectations, and whether such modulation of WM is stronger for the lexical than for the phonetic task, as the phonetic task might be more automatic and not require as many resources.

1.2.2. Autistic traits

Autism Spectrum Disorder (ASD) is associated, among other things, with detail-focus processing style, and reduced attention to social dimensions [21]. Both tendencies could reduce susceptibility to the influence of expectations as it requires integration of the speech input in its social context. Autistic traits exist on a spectrum, and so individuals in the typical population also exhibit these traits to a higher or lower degree. Furthermore, previous research suggests that the degree of autistic traits even in the typical population can predict linguistic performance, such as compensation for co-articulation during speech perception [22-23]. We test whether autistic traits lead to reduced influence of expectations, and whether this occurs similarly at the phonetic and lexical levels.

1.2.3. Attitudes

Variations in pronunciation often elicit strong attitudes, in part due to their associations with demographic parameters. Such attitudes could influence the way the speech is processed. For example, it has been proposed that encoding and representation of speech tokens depend on their prestige, with more prestigious token given greater weight [24]. Therefore, attitudes might influence the degree to which expectations influence processing, especially if the integration of social information with the speech is effortful, as individuals with negative attitudes might devote fewer resources for processing. Additionally, attitudes might correlate with expectations, thus influencing the availability and stereotypicality of expectations, and consequently, their influence. We therefore test whether attitudes moderate the effect of expectations and processing, and whether they do so similarly at the phonetic and lexical levels.

2. Study

The current study has several goals. First, it aims to test performance on acoustic and lexical tasks within the same experiment and thus examine whether the two are comparable. It further tests the prediction that expectations lead to exaggeration in the phonetic task but compensation in the lexical task. Finally, it tests whether the effect of expectations on speech perception is modulated by cognitive and social individual differences, and if so, whether performance at the acoustic and lexical levels is similarly sensitive to the same individual differences.

In order to manipulate the English varieties participants expect the speaker to produce, we present listeners with speech of speakers that are presented as residents either of South Carolina (Southern variety) or of Ohio (General American variety, henceforth GA). We exploit the fact that several vowel phonemes are pronounced differently in Southern compared to GA varieties. Specifically, /æ/ resembles [ɛ], /ɑɪ/ resembles [ɑ], /ɔɪ/ resembles [ɔʊ], and /eɪ/ resembles [ɑɪ] [25]. Participants are residents of North Carolina, where GA and Southern norms coexist. They perform a vowel matching task and a lexical interpretation task using these target vowels. We additionally test participants' WM, autistic traits, and implicit attitudes.

2.1. Methods

2.1.1. Participants

Seventy-three residents of Raleigh, NC, participated in the study for pay (F=32). Their ages ranged from 19-72 (M:33, SD=18.2). One additional participant took part but was excluded since she was not originally from North Carolina. Due to technical failure, the WM measure is missing for two participants and the implicit attitude measure for another two.

2.1.2. Stimuli

Word interpretation. For each of the contrasts /æ/-/ɛ/, /ɑɪ/-/ɑ/, /ɔɪ/-/ɔʊ/, and /eɪ/- /ɑɪ/, we selected six GA minimal pairs with monosyllabic words, such that a Southern pronunciation of one word (e.g., *sad*) might be confused with the second word in GA pronunciation (e.g., *said*). For each minimal pair we constructed a sentence that could fit with both meanings but fits better with the meaning that is in line with GA pronunciation (e.g., *Jill was sad/said to be away from her hometown*). All sentences were normed beforehand to ensure their fit and bias. Forty additional filler sentences were constructed, none of which included the target vowels. One speaker who speaks a Southern variety characteristic of Central North Carolina recorded all sentences. The target sentences were presented to her with the first member of the minimal pair (e.g., *sad*). For each sentence, a question was generated to probe how participants interpreted the potentially ambiguous word (e.g., *Do we know how Jill feels about being away from her hometown?*).

Vowel matching. For each of the GA vowels /æ/, /ɑɪ/, /ɔɪ/, and /eɪ/, we selected 15 monosyllabic words such that replacing the vowel by its Southern correspondent does not yield another word in GA. A resident of Raleigh whose speech includes features of the Southern variety recorded all words, and we selected two tokens for each of them. The speaker's accent falls in the middle of the range of the speakers in the Raleigh corpus collected by the second author. We also created four synthetic vowel continua, each ranging from one of the four GA vowels to its Southern correspondent. Each continuum had five equally spaced steps, all 300ms long. The end poles had formant frequencies that are a bit more exaggerated than the end poles of the recorded speaker for these vowels.

Working Memory task. We used the standard Operation Span to measure WM [26]. It is comprised of equations to solve and letters to memorize.

Autism-spectrum Quotient (AQ). We used the standard AQ questionnaire, [27]. It is comprised of 50 questions targeting several dimensions of ASD, including attention to detail and social skills.

Implicit Bias task. We created a Single Category Implicit Association Test. The task included eight positive words (e.g., *joy*), eight negative words (e.g., *horrible*), and eight words associated with Southern USA (*Alabama*, *Mississippi*, *South Carolina*, *Louisiana*, *barbecue*, *bluegrass*, *collards*, *cornbread*).

2.1.3. Procedure

To mask the goal of the experiment, we presented the tasks to all participants in the following order, with the more explicit tasks appearing at the end: word interpretation task, WM measure, AQ questionnaire, vowel matching task, and single category IAT. The experiment took less than an hour.

Word interpretation. Participants were told that the experiment was done in collaboration with a lab in Ohio / South Carolina (counterbalanced across participants). On each trial participants listened to a sentence. Five seconds after sentence onset, an open-ended question appeared on the screen, and participants typed their responses. The task opened with two practice trials. All sentences appeared in random order.

Working Memory task. The standard Operation Span procedure was followed [26]. Participants evaluated equations while memorizing sets of letters.

Autism-spectrum Quotient. Participants answered the questions of the questionnaire [27], in one fixed order.

Vowel matching. Participants were told that the lab was developing a speech synthesizer and needed help adjusting it to speakers from different regions of the country. Participants were then told that they would listen to a speaker from Ohio / South Carolina (counterbalanced across participants). On each trial, participants heard the target word, and 1400ms after its onset, they heard the five vowels on the relevant continuum in succession with an inter-stimulus interval of 300ms. Each vowel on the continuum appeared with a number indicating its position on the continuum (1-5). After the entire continuum was presented, participants indicated which vowel was closest to the vowel in the target word by typing a number between 1 and 5. Words were presented in random order. The task started with four practice trials.

Implicit Bias task. Participants categorized words into positive, negative, and related to South USA. On each trial, they saw a word in the center of the screen and classified it by pressing one of two keys as quickly as possible. If participants made an error, a screen with the word 'incorrect' in red font was displayed for 500ms. In one block, classification as South USA was done with the same key as classification of positive words, and in the other block with the same key as negative words. The order of these blocks was counterbalanced across participants. Each block started with 28 practice trials, followed by 72 test trials. In order to avoid bias due to mostly using one key in each block, words from the category that did not share a key with another category appeared 1.5 times more frequently than words in each of the other two categories. Implicit bias was measured as the RT in the block in which South USA was grouped with positive words divided by the RT in the block in which it was grouped with negative words. Thus, higher values indicate prejudice against southern USA.

2.2. Results

All individual differences exhibited wide variation: WM scores ranged from 32 to 74 (on a scale from 0 to 75) with an average of 58 (SD=9.79). AQ scores ranged from 8 to 39 (on a scale from 0 to 50) with a mean of 19.65 (SD=6.58). IAT scores ranged from 0.69 to 1.39 with a mean of 0.92 (SD=0.12) reflecting the fact that most participants had a positive attitude towards southern USA.

2.2.1. Word interpretation

One participant was excluded because she reported to have used the buttons incorrectly. Responses that reflected interpretation in line with GA pronunciation were coded as 1, and responses that reflected interpretation in line with Southern pronunciation were coded as 0. We excluded 104 responses (5.8%; 59 in GA condition, 45 in Southern condition) that could not be unambiguously categorized into either category. To test whether participants were influenced by the identity of the speaker, and whether this influence was modulated by individual differences, we ran a logistic mixed effects regression with Identity (Ohio vs South Carolina), WM (centered), IAT (centered), and AQ (centered) as fixed effects, including the interactions of each individual difference measure with Identity. The random structure included Participants and Items. Slopes were not included as their inclusion led to convergence errors or singular fit.

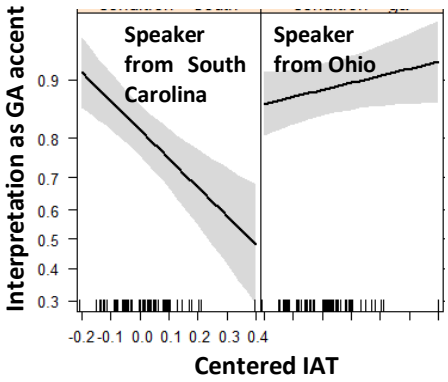

Figure 1: *Lexical interpretation as dependent on speaker identity and listener's implicit attitude towards South USA. Shaded bands indicate Standard Errors.*

Results showed an effect of Identity (β=-0.55, SE=0.24, z=-2.31, p<0.03), such that participants were more likely to interpret the target words in line with a GA pronunciation when the speaker was said to be from Ohio than when she was said to be from South Carolina. The direction of this result indicates an effect of compensation. Furthermore, results showed an effect of IAT at the reference level (South; β=-3.95, SE=1.84, z=-2.15, p<0.03) and an interaction of Identity with IAT (β=-4.92, SE=2.19, z=-2.25, p<0.03). As Figure 1 illustrates, when the speaker was presented as being from South Carolina, participants who are less positive about South USA were more likely to compensate and interpret the target word in line with Southern pronunciation. Attitudes did not influence performance when the speaker was presented as a resident of Ohio. No other main effect or interaction came close to significance.

The direction of the effect of Identity supports our prediction that at the lexical level expectations lead to compensation. The effect of implicit attitudes is more surprising. One might think that less positive attitudes would lead to decreased effort in processing, and therefore potentially reduce the influence of expectations, yet we find the opposite. One reason might be that those with a less positive attitude towards South USA might have a more stereotypical view of Southern speakers, whereas those with more positive attitudes might have more nuanced expectations. Those with more positive attitudes might be better aware of the great variation in accents that exists in South USA, and would therefore not necessarily expect any southern speaker to produce words with typical southern accent. In contrast, those with less positive attitude might be less aware of the variation and hold strong categorical expectations. Indeed, bias often leads to individuation of ingroup members but homogenization of outgroup members [29]. Note also that attitude towards South USA only influences interpretation of the southern speaker. When the speaker is presented as a resident of Ohio, all listeners interpret her speech in line with GA pronunciation, as attitudes towards South USA should not play a role when the speaker is not from that region.

2.2.2. Vowel matching task

To test the effect of expectations on vowel matching and its modulation by individual differences, we first averaged participants' responses for each continuum, and then normalized these responses across participants for each continuum. We then ran a regression on participants' normalized responses with Identity (Ohio vs South Carolina), Continuum, WM (centered), IAT (centered), and AQ (centered), as well as the interaction of each individual difference measure with Identity as predictors.

Results revealed an effect of WM at the reference level (South; β=0.03, SE=0.01, t=4.15, p<0.001) and an interaction of Identity and WM (β=-0.04, SE=0.01, t=-2.69, p<0.01). As Figure 2 illustrates, higher WM was associated with a greater effect of exaggeration when the speaker was presented as a South Carolina resident, but did not influence perception when the speaker was presented as a resident of Ohio. No other main effect or interaction came close to significance.

The results of this task did not reveal the predicted effect of exaggeration, as there was no main effect of Identity. At the same time, the observed interaction is in line with the hypothesis that the influence of expectations relies on WM. It is unclear, however, why such integration would be effortful at the phonetic level, but not at the lexical level, and similarly, why the effect of attitudes was restricted to the lexical task.

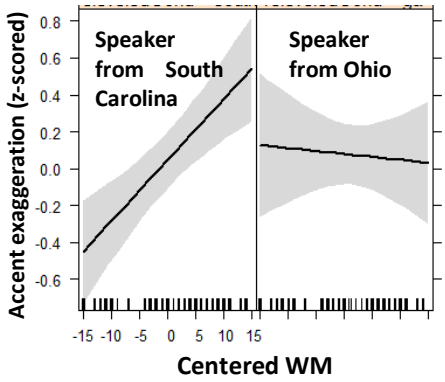

Figure 2: *Exaggeration of produced accent as dependent on speaker identity and listener's WM. Shaded bands indicate Standard Errors.*

3. Conclusion

Speech perception is influenced by the expectations that listeners have of speakers. Previous literature has demonstrated this effect at both the phonetic and lexical levels. Interestingly, effects in any direction at the two levels were considered equivalent and assumed to reflect the same mechanism. Here we show that while expectations could influence perception at both these levels, the two effects are not equivalent. At the phonetic level, while there was no main effect of Identity, the interaction with WM suggests that expectations lead to exaggeration. At the lexical level, in contrast, results showed an effect of Identity in the direction of compensation. Furthermore, we demonstrate that the degree to which expectations influence perception depends on listeners' cognitive skills and attitudes, with WM influencing performance at the phonetic level and implicit attitude influencing performance at the lexical level. The finding that individual differences play different roles at the phonetic and lexical levels supports the proposal that the influence of expectations at these levels is driven by two distinct mechanisms. As phonemes are embedded within words, future research should not only further develop the proposed mechanism by which expectations influence performance at the different levels, but also examine the relative weight of each during processing. For example, the level of ambiguity at each level might influence their relative weight. Importantly, future studies should avoid treating the phonetic and lexical paradigms as equivalent and instead either choose the paradigm most appropriate to the question, or compare performance across both paradigms.

4. Acknowledgements

Thanks to Megan Risdal, Eric Wilbanks, and Christopher Carignan for running participants. This study was supported by ANR-17-EURE-0017, ANR-17-CE28-0007-01, and the North Carolina State University Department of English.

5. References

[1] W. Labov, S. Ash, and C. Boberg. *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter, 2008.

[2] W. Labov. *Sociolinguistic patterns*. No. 4. University of Pennsylvania Press, 1972.

[3] J. Rickford. "Ethnicity as a sociolinguistic boundary," *American Speech*, vol 60, pp. 99–125, 1985.

[4] P. Eckert. *Jocks and burnouts: Social categories and identity in the high school*. Teachers college press, 1989.

[5] W. Labov. *Principles of linguistic change, volume 3: Cognitive and cultural factors*. Vol. 36. John Wiley & Sons, 2011.

[6] M. Brunelle and S. Jannedy. "Social effects on the perception of Vietnamese tones," *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 1461-1464, 2007.

[7] K. Drager. "Speaker age and vowel perception." *Language and Speech*, vol 54, no.1, pp. 99-121, 2011.

[8] J. Hay, A. Nolan, and K. Drager. "From fush to feesh: Exemplar priming in speech perception." *The linguistic review,* vol 23, no. 3, pp. 351-379, 2006.

[9] J. Hay, P. Warren, and K. Drager. "Factors influencing speech perception in the context of a merger-in-progress." *Journal of phonetics,* vol 34, no. 4, pp. 458-484, 2006.

[10] S. Jannedy, M. Weirich, and J. Brunner. "The Effect of Inferences on the Perceptual Categorization of Berlin German Fricatives." *Proceedings of the 17th International Congress of Phonetic Sciences, pp. 962-965,* 2011.

[11] K. Johnson, E. A. Strand, and M. D'Imperio. "Auditory–visual integration of talker gender in vowel perception." *Journal of phonetics,* vol 27, no. 4, pp. 359-384, 1999.

[12] C. Koops, E. Gentry, and A. Pantos. "The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas." *University of Pennsylvania Working Papers in Linguistics,* vol 14, no. 2, aritcle. 12, 2008.

[13] N. Niedzielski. "The effect of social information on the perception of sociolinguistic variables." *Journal of language and social psychology,* vol 18, no. 1, pp. 62-85, 1999.

[14] D. L. Rubin. "Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants." *Research in Higher education,* vol 33, no. 4, pp. 511-531, 1992.

[15] L. Staum Casasanto. "Does social information influence sentence processing?." *Proceedings of the Annual Meeting of the Cognitive Science Society*. vol. 30. No. 30, pp. 799-804, 2008.

[16] S. D. Goldinger. "Echoes of echoes? An episodic theory of lexical access." *Psychological review,* vol 105, no. 2, pp. 251-279, 1998.

[17] K. Johnson. "Speech perception without speaker normalization: An exemplar model." *Talker variability in speech processing,* pp. 145-165, 1997.

[18] M.A Just, and P.A. Carpenter. "A capacity theory of comprehension: individual differences in working memory". *Psychological Review,* vol 99, pp. 122–149, 1992.

[19] M.J. Traxler, R.S. Williams, S.A. Blozis, and R.K. Morris. "Working memory, animacy, and verb class in the processing of relative clauses", *Journal of Memory and Language,* vol 53, pp. 204–224, 2005.

[20] S. Lev-Ari. "Comprehending non-native speakers: theory and evidence for adjustment in manner of processing." *Frontiers in psychology,* vol 5, pp. 1546, 2015.

[21] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[22] A. C. L, Yu. "Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change." *PloS one,* vol 5, no. 8, pp. e11950, 2010.

[23] J. Mielke, K. Nielsen, and L. V. Magloughlin. "Phonetic imitation by individuals with Autism Spectrum Disorders: Investigating the role of procedural and declarative memory." *Proceedings of Meetings on Acoustics ICA2013*. Vol. 19. No. 1. ASA, 2013.

[24] M. Sumner, S.K. Kim, E. King & K.B. McGowan. "The socially weighted encoding of spoken words: A dual-route approach to speech perception." *Frontiers in Psychology,* vol 4 pp. 1015, 2014.

[25] R. Dodsworth and M. Kohn. "Urban rejection of the vernacular: The SVS undone." *Language Variation and Change,* vol 24, no. 2, pp. 221-245, 2012.

[26] N. Unsworth, R.P. Heitz, J.C. Schrock, and R.W. Engle. "An automated version of the operation span task," *Behavioral Research Methods,* vol 37, pp. 498–505, 2005.

[27] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley. "The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians." *Journal of autism and developmental disorders,* vol 31, no. 1, pp. 5-17, 2001.

[28] A. Karpinski, and R. B. Steinman. "The single category implicit association test as a measure of implicit social cognition." *Journal of personality and social psychology,* vol 91, no. 1, pp. 16-32, 2006.

[29] K. Hugenberg, S.G. Young, M.J. Bernstein, and D.F. Sacco. "The categorization-individuation model: an integrative account of the other-race recognition deficit," *Psychological Review,* vol 117, no. 4, pp. 1168-1187, 2010.